

Fri Jun 4 10:08:27 1999

US08945574.pep

Sequence 1, Application US/08945574
GENERAL INFORMATION:
APPLICANT: Lenting, Hermanus Bernardus Maria
APPLICANT: Van Beckhoven, Rudolf Franciscus Wilhelmus Cornelis
APPLICANT: Maurer, Karl-Heinz
APPLICANT: Kottwitz, Beatrix
APPLICANT: Weiss, Albrecht
APPLICANT: Van Solingen, Pieter
TITLE OF INVENTION: Detergents Comprising Cellulases
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Henkel Corporation
STREET: 140 Germantown Pike, Suite 150
CITY: Plymouth Meeting
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19462
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" diskette
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: MS Word 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/945,574
FILING DATE: unavailable
CLASSIFICATION: C11D 3/386
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP96/01755
FILING DATE: 26 Apr. 1996
APPLICATION NUMBER: EP 95201115.3
FILING DATE: 28 Apr. 1995
APPLICATION NUMBER: U.S. 614,115
FILING DATE: 12 Mar. 1996
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Glenn E. J.
REGISTRATION NUMBER: 33,539
REFERENCE/DOCKET NUMBER: H 1920 PCT/US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 832-2228
TELEFAX: (610) 941-6067
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 457 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-945-574-1
MKRTTIFAVLLMTLALFSICNTTAAADYSVVEERGLSISNGELVNERGEYOLKGMSSHGLQWYGFV
NYESMKLWDDMGITVPEAAVNTSSGGYIDPSVKEKVEVAIDLGIVLIDHLLSDNDPHIYKEE
AKDFDEMSLGYDPNVYVILIANEPNGSYTDNCKIKFYAEVPIVRNDNNVIVGTGTWSDVHH
AANQLADPNVYAHFTAGTGNLQADQVDYALQOGAAIFVSENGTSAAATGGGVFLDEAQVWIDFDE
RNLNANWLSLTHKDESSAALMPGANPGTGGWTEAELSPSGTFVREKIRIESASIPSPDPTPPSGPDE
POPTPPSPGEPYPAWDSNQIYTNIEVYHNGOLWQAQWNTQNOEPDYGWPEPLNI
PGEYPAWDSNQIYTNIEVYHNGOLWQAQWNTQNOEPDYGWPEPLNI

Sequence 2, Application US/08945574
GENERAL INFORMATION:
APPLICANT: Lenting, Hermanus Bernardus Maria
APPLICANT: Van Beckhoven, Rudolf Franciscus Wilhelmus Cornelis
APPLICANT: Maurer, Karl-Heinz
APPLICANT: Kottwitz, Beatrix
APPLICANT: Weiss, Albrecht
APPLICANT: Van Solingen, Pieter
TITLE OF INVENTION: Detergents Comprising Cellulases
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Henkel Corporation
STREET: 140 Germantown Pike, Suite 150
CITY: Plymouth Meeting
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19462
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" diskette
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: MS Word 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/945,574
FILING DATE: unavailable
CLASSIFICATION: C11D 3/386
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP96/01755
FILING DATE: 26 Apr. 1996
APPLICATION NUMBER: EP 95201115.3
FILING DATE: 28 Apr. 1995
APPLICATION NUMBER: U.S. 614,115
FILING DATE: 12 Mar. 1996
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Glenn E. J.
REGISTRATION NUMBER: 33,539
REFERENCE/DOCKET NUMBER: H 1920 PCT/US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 832-2228
TELEFAX: (610) 941-6067
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 574 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-945-574-2
MKHKSVMILAVLVVYFVPAVSSANEDVKTLIDIOSYVRDMQPCWNLGNTFDVAGQDETAMGNPRVTR
LIERIADGKTSIRIPVTEWRIGGAPYIDPQFLNRVDEYVOWALEEDLYVNLHDSWLIYEMEH
NYNGVAKTSVWEGSLNHFXYDTYTKLWFSVNEKFSQNGEIRHNHALLDLNTVFEIVRQSGGN
DIRPLVLTMETATQPLNLNLYOTIDKLDDPNLIATVHYGFWPFSVNIAGYTRFEEDSKREIIFEDR
VHFTFVARGIPVYVGLGFLGDKHTGVIOQGEKLAFFEYLHILNEDITHMLWNGOHFNHFTYKVD
BELFDMLRASGRSSVAESNFYILKQGDRIADATVTLQHGNETJGLQANGRLTFQGDYELNGERLIV
KARVLSAAGSGTLATNGWTAERNAGADHFRVNTYRTPVLSQSTQGHVSNFSPASFGNSLATMEAVY
VDGNGRQPDWTSFREFGYAFSPSYDTHIRLTAFFREVROGEVRLTFHFWSGEIVNYYTIKNGNQV7G
IAQTINSNNKNNKKI

MPREH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

mparch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 4 09:55:50 1999; MasPar time 14.37 Seconds

Tabular output not generated. 918.961 Million cell updates/sec

Title: >US-08-945-574-1
Description: (1-467) from US08945574.pep
Perfect Score: 3419

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 50.573; Variance 100.476; scale 0.503

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2828	82.7	409	1	GUN2_BACS4 ENDOGLUCANASE B (EC 3.1.1.1)	0.00e+00
2	2803	82.0	488	1	GUN1_BACS4 ENDOGLUCANASE A (EC 3.1.1.1)	0.00e+00
3	1664	48.7	505	1	GUNV_ERWCA ENDOGLUCANASE V PRECURSOR	0.00e+00
4	1660	48.6	499	1	GUN1_BACSU ENDOGLUCANASE PRECURSOR	0.00e+00
5	1659	48.5	444	1	GUNN_ERWCA ENDOGLUCANASE N PRECURSOR	0.00e+00
6	1641	48.0	499	1	GUN2_BACSU ENDOGLUCANASE PRECURSOR	0.00e+00
7	1616	47.3	504	1	GUNW_ERWCA ENDOGLUCANASE V1 PRECURSOR	2.51e-302
8	1598	46.7	499	1	GUN3_BACSU ENDOGLUCANASE PRECURSOR	1.64e-298
9	1491	43.6	448	1	GUN_CLOBR ENDOGLUCANASE PRECURSOR	7.41e-276
10	884	25.9	429	1	GUNA_BUTFI ENDOGLUCANASE A PRECURSOR	1.04e-148
11	816	23.9	426	1	GUN2_ERWCA ENDOGLUCANASE Z PRECURSOR	1.08e-134
12	810	23.7	312	1	GUN4_RUMAL ENDOGLUCANASE IV (EC 3.1.1.1)	1.84e-133
13	805	23.5	466	1	GUN5_THEFU ENDOGLUCANASE E-5 PRECURSOR	1.96e-132
14	739	21.6	800	1	GUN_BACSI ENDOGLUCANASE PRECURSOR	6.52e-119
15	727	21.3	459	1	GUNA_STRLI ENDOGLUCANASE CELA PRECURSOR	1.84e-116
16	690	20.2	941	1	GUN_BACSG ENDOGLUCANASE PRECURSOR	6.29e-109
17	613	17.9	825	1	GUN3_BACS4 ENDOGLUCANASE C PRECURSOR	2.39e-93
18	186	5.4	848	1	GUND_CLOCE ENDOGLUCANASE D PRECURSOR	1.40e-12
19	179	5.2	900	1	GUNH_CLOTHM ENDOGLUCANASE H PRECURSOR	1.88e-11
20	175	5.1	820	1	CHIA_ALRSM CHITININASE A PRECURSOR	8.17e-11
21	139	4.1	826	1	SSP2_PLAYO SPOROZOITE SURFACE PRO	2.39e-05
22	128	3.7	897	1	YHEB_ECOLI HYPOTHETICAL 97.1 KD P	8.45e-04
23	122	3.6	699	1	VGLG_HSV2H GLYCOPROTEIN G.	5.51e-03

24	124	3.6	1331	1	MANB_CALSA BETA-MANNANASE / ENDOS	2.97e-03
25	120	3.5	316	1	UNC_PVKA URACIL-DNA GLYCOSYLASE	1.02e-02
26	116	3.4	339	1	CSP_PLABE CIRCUMSPOROZOITE PROTE	3.40e-02
27	115	3.4	347	1	CSP_PLABA CIRCUMSPOROZOITE PROTE	4.57e-02
28	116	3.4	353	1	CA26_HUMAN COLLAGEN ALPHA 2(VI) C	3.40e-02
29	117	3.4	449	1	GUNA_CELFI ENDOGLUCANASE A PRECUR	3.40e-02
30	117	3.4	825	1	5E5_RAT 5E5 ANTIGEN	2.52e-02
31	112	3.3	456	1	GUNA_MICBI ENDOGLUCANASE A PRECUR	1.10e-01
32	114	3.3	487	1	EBN2_EBV EBNA-2 NUCLEAR PROTEIN	6.14e-02
33	112	3.3	658	1	ENDOGLUCANASE 3 PRECUR	1.10e-01
34	108	3.2	143	1	PAR1_TRYBB PROCYCLIC FORM SPECIFI	3.49e-01
35	108	3.2	145	1	PARC_TRYBB PROCYCLIC FORM SPECIFI	3.49e-01
36	108	3.2	185	1	ADML_HUMAN ADRENOMEDULLIN PRECURS	3.49e-01
37	108	3.2	295	1	LEG3_CANFA GALECTIN-3 (GALACTOSE-	3.49e-01
38	108	3.2	315	1	UNG_PRIVIF URACIL-DNA GLYCOSYLASE	3.49e-01
39	108	3.2	349	1	FLAV_SOLTU FLAVONOL SYNTHASE (EC	3.49e-01
40	109	3.2	484	1	GUX_CELFI EXOGLUCANASE PRECURSOR	2.62e-01
41	111	3.2	917	1	SVI_STAUI ISOLEUCYL-TRNA SYNTHET	1.48e-01
42	111	3.2	1029	1	CA26_MOUSE COLLAGEN ALPHA 2(VI) C	1.48e-01
43	110	3.2	1664	1	SLPI_CLOTH CELL SURFACE GLYCOPROT	1.97e-01
44	106	3.1	459	1	G33_RAT GENE 33 POLYPEPTIDE	6.12e-01
45	107	3.1	1018	1	FNBA_STAUI FIBRONECTIN-BINDING PR	4.63e-01

ALIGNMENTS

RESULT 1
ID GUN2_BACS4 STANDARD; PRT; 409 AA.
AC P06565;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE B (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE)
DE (CLONE PNK2).
GN CELB.
OS BACILLUS SP. (STRAIN N-4).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87056924.
RA FUKUMORI F., SASHIHARA N., KUDO T., HORIKOSHI K.;
RT "Nucleotide sequences of two cellulase genes from an alkalophilic
Bacillus sp. strain N-4 and their strong homology";
RL J. BACTERIOL. 168:479-485(1986).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M14729; G142656; -
CC PIR; B25156; B25156.
CC PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
CC PFAM; PF00150; cellulase; 1
CC CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE.
KW ACT_SITE 165 165 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 254 254 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 409 AA; 45690 MW; C77CF2F8 CRC32;

Query Match 82.7%; Score 2828; DB 1; Length 409;
Best Local Similarity 94.6%; Pred. No. 0.00e+00;
Matches 387; Conservative 12; Mismatches 6; Indels 4; Gaps 2;

Db 1 MKKITTIFVLLMTALFIIGNTTAADDYVVEEHGQLSISNGELVNDRGEVQLKGMS 60
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Query Match 82.0%; Score 2803; DB 1; Length 488;
Best Local Similarity 84.0%; Pred. No. 0.00e+00;
Matches 414; Conservative 30; Mismatches 18; Indels 31; Gaps 9;
Db 1 MKKLTTFIVFTIAL-LF-VGNSTSANNGSWEQNGOLSIQGLVNEHGDVPOLKGMSS 58

QY		: :: : : : : : : : : : : : : : : : : : : : : : : :	
1	MKKITTFVLLMTLALFISGNTTAAADDSVVEEHQQLSINGELVNERGEQVLKGMSS	60	
D59	HGLWYGQFNVDYSIKWLDDMGITVFRAMYTSSGGYTDPSVKKVKAEAVEAADIIGI	118	
	: :		
QY	61 HGLWYGQFNYESMKWLDDMGITVFRAMYTSSGGYIDDPVKKVKETVEAADIGI	120	
	: :		
D619	VVIIDHILSDNDPNITYKEAKAEFFDEMSELYGDYPNVIVYEIANEPNGHNVRWDSHIKPY	178	
QY	121 VVIIDHILSDNDPNITYKEAKAEFFDEMSELYGDYPNVIVYEIANEPNGSDVTWNQIKPY	180	
	: : :~ :		
D619	AEEVIPVIRANDPNNIVICTATWSQDVHEADNQLDDPNMYAFHYAGTHGOQLRNQY	238	
	: :		
QY	181 AEEVIPVIRDNDPNNIVIGTGTSQDVHHAADNLADPNMYAFHYAGTHGONLRDQV	240	
	: :		
D6239	DYALSRAAIFVSEWGTSAAATGGGVFLDEAQWIDFMDERNLSPANWLSLTHKDESSAAL	298	
QY	241 DYALDQGAALFVSEWGTSAAATGGGVFLDEAQWIDFMDERNLSPANWLSLTHKDESSAAL	300	
	: :		
D6299	MPGANPTGGWTAELSFGSAFVREKIREASIPSPDTPPSDP - --DPGEPPDTPPSDPG	355	
QY	301 MPGANPTGGWTEALSFGSFVREKIREASIPSPDTPPSDPGEPPDTPPSDPG	360	
	: :		
D6356	EYPAMDNQIYTNIEIVYHNGLMQAOKWWTQNQBPGANQGWPBPLGDAPSPSPDPPPS	415	
QY	361 EYPAMDNQIYTNIEIVYHNGLMQAOKWWTQNQBPG-DPYGPWEPL-K---SDP-D----S	410	
	: :		
D6416	EPEPDCEPDGPDGEDPTPPSDPGEYPAWMDPTQIYTNIEIVYHNGLMQAOKWWTQNO	475	
QY	411 -----GEPDTPPSDPGEYPAWMSNQIYTNIEIVYHNGLMQAOKWWTQNO	454	
	: :		
D6476	EPGYPGYPEPLN 488		
	: :		
QY	455 EPGDYPGYPEPLN 467		
	: :		
RESULT	3		
ID	GUNV_ERWCA STANDARD; PRT; 505 AA.		
AC	Q47096;		
DT	01-NOV-1997 (REL. 35, CREATED)		
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)		
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)		
DE	ENDOGLUCANASE V PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE V)		
DE	(CELLULOSE V).		
GN	CELV.		
OS	ERWINIA CAROTOVORA.		
OC	BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;		
RC	ERWINIA.		
NC	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=SCRI3;		
RX	MEDLINE; 94067016.		
RA	COOPER V.J.C., SALMOND G.P.C.;		
RT	"Molecular analysis of the major cellulase (Celv) of Erwinia		
RT	carotovora: evidence for an evolutionary 'mix-and-match' of enzyme		
RT	domains";		
RL	MOL. GEN. GENET. 241:341-350(1993).		
CC	-1- FUNCTION: ENDOGLUCANASE WITH SOME EXOGLUCANASE ACTIVITY. THE PH		
CC	OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGREES		
CC	CELSIUS.		
CC	-1- CATALYTIC ACTIVITY: ENDOTHERMOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC		
CC	LINKAGES IN CELLULOSE.		
CC	-1- SUBCELLULAR LOCATION: SECRETED.		
CC	-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL		
CC	HYDROLASES).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-		
CC	tion between the Swiss Institute of Bioinformatics and the EMBL Outstat-		
CC	the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no		
CC	modified and this statement is not removed. Usage by and for commer-		
CC	c entities requires a license agreement (See http://www.isb-sib.ch/anno-		
CC	or send an email to licenses@isb-sib.ch).		

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC	ENBL; M16185; G143008; ALT_INIT.
DR	PIR; A26874; A26874.
DR	PROSITE; PS00659; GLYCOSYL-HYDROL_F5; 1.
DR	PFAM; PF00150; cellulase; 1.
DR	PFAM; PF00942; CBD_3; 1.
DR	HSP; Q06851; INBC.
KW	CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
FT	SIGNAL 1 29
FT	CHAIN 30 499
FT	ACT_SITE 169 169
FT	ACT_SITE 257 257
FT	DOMAIN 350 499
SQ	SEQUENCE 499 AA; 55187 MW; C38F2C29 CRC32;

Query Match 48.6%; Score 1660; DB 1; Length 499;
 Best Local Similarity 62.7%; Pred. No. 0.00e+00;
 Matches 222; Conservative 63; Mismatches 63; Indels 6; Gaps 6;

Db	1	MKRISIFITCLLTAVITMGLLSPASAAGTKPVAKNGQLSKGTQLVNRDKAVOLK	60
QY	1	MKKTITFAV-LMT-LALFSI-GNTTAADYSV-VEEHGQLSISNGELVNERGEQVOLK	56
Db	61	GISSHGLQWYGFYKNDKSLKLRDWDGTTVPRAAMYADGGYIDNPVKNVKAVERAAK	120
QY	57	GMSSHGLQWYGFYVYSKWLRRDWTGTTVPRAAMYSGGYIDNPVKNVKAVERAAI	116
Db	121	ELGIYVIDHMLNDGNPNQNKAKAEFFKEMSSLYGNTPNVIYEIANEPNG-DYNMKRD	179
QY	117	DLGIYVIDHMLSDNDPNYKKEAKDFFDEMSELYGDPNVIYEIANEPNGSDVTWDNQ	176
Db	180	IKPYAEVIVIRKNDPNIIIVGTQWSQVNDAAADDOLKANDVMYALHFIYAGTHGSL	239
QY	177	IKPYAEVIVIRKNDPNIIIVGTQWSQVNDAAADDOLKANDVMYALHFIYAGTHGSL	236
Db	240	RDKANYALSKAPITFVETWGTSDASNGGVFLDQSRWLNLDKSNISWANNISDKOES	299
QY	237	RQVDYALDQGAALFVSEWGTSAITDGGVFLDQAWIDFMDERNLSWANWSLTHKDES	296
Db	300	SSALPKGACGWLPLDLTSLATFVREINRIGTKDSTKDVPEPAQDNPTQEK	353
QY	297	SAALMPGANPTGGWTEALSPGTFVREKIREASISPPSDP-TTPSDPGPDGP	349

RESULT 5

ID	GUNN_ERWCA	STANDARD;	PR1;	444	AA.
AC	Q59394;				
DT	01-NOV-1997 (REL. 35, CREATED)				
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)				
DE	ENDOGLUCANASE N PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE N (CELLULASE N).				
GN	CELN.				
OS	ERWINIA CAROTOVORA				
OC	BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;				
OC	ERWINIA.				
RN	[1]				
RC	SEQUENCE FROM N.A.				
RC	STRAIN-ATROSEPTICA FCBR C18;				
RX	MEDLINE; 98299944.				
RA	OLSEN O., THOMSEN K.K., WEBER J., DUUS J.O., SVENDSEN I., WEGENER C.,				
RA	VON WETTINGEN D.;				
RT	"Transplanting two unique beta-glucanase catalytic activities into				
RT	one multienzyme, which forms glucose."				
RL	BIOTECHNOLOGY 14:71-76(1996).				
CC	-1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC				
CC	LINKAGES IN CELLULOSE.				
CC	-1- SUBCELLULAR LOCATION: SECRETED.				
CC	-1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL				
CC	HYDROLASES).				

RA LINDAHL V., AA K., TRONSMO A. ;
RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from *Bacillus*
RT subtilis CK-2";
RL ANTONIE VAN LEEUWENHOEK 66:327-332(1994) .
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA ROSE M., ENTIAN K.;

RL SUBMITTED (JUN-1996) TO EMBL/GENEBANK/DBJ DATA BANKS.
 [4]
 RN SEQUENCE OF 30-45.
 RP
 RC STRAIN-CK-2;
 RX
 RA AA K., FLENGSRUD R., LINDAHL V., TRONSMO A.;
 RT "Characterization of production and enzyme properties of an
 RT endo-beta-1,4-glucanase from *Bacillus subtilis* CK-2 isolated from
 RT compost soil.";
 RL ANTONIE VAN LEEUWENHOEK 66:319-326(1994).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC

LINKAGES IN CELLULOSE.
-!- SIMILARITY; BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
HYDROLASES).

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

DR	EMBL; Z29076; G509267; -.
DR	EMBL; X04689; G39824; -.
DR	EMBL; X67044; G39777; -.
DR	EMBL; Z73234; E249661; ALT INIT.

DR PIR; A26114; A26114.
DR DR SUBTILIST; BG10437; BGIC.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PFAM; PF00150; cellulase; 1.

DR PFAM; PF00942; CBD_3; 1.
DR HSSP; Q06851; 1NBC.
KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
FT SIGNAL 1 29

FT	CHAIN	30	ENDOGLUCANASE.
FT	ACT_SITE	169	PROTON DONOR (BY SIMILARITY).
FT	ACT_SITE	257	NUCLEOPHILE (BY SIMILARITY).
FT	DOMAIN	350	CELLULOSE-BINDING (BY SIMILARITY).

FT	CONFLICT	283	283	S -> N (IN REF. 2).
SQ	SEQUENCE	499 AA;	55287 MW;	1DE2AA90 CRC32;
	Query Match		48.0%;	Score 1641; DB 1; Length 499;

Best Local Similarity 62.7%; Pred. No. 0.00e+00;
Matches 222; Conservative 60; Mismatches 66; Indels 6; Gaps 6;

Qy 1 MKKITTFAV-LLMTL-ALFS-IGNTTAADDYSV-VEEHGQLSINGELVNERGEQYQLK 56

Dd 61 GISSHGLOWIGEYVKNQSLKWLRRDWDGITTFRAMYYTADGGYIDNP5YKNVKAFAEAAK 120

Qy 57 GWSHGLQWYQFYNESMKWLWDGITTFRAMYTSSGGYIDDPSPKEVKVETVAAL 116
| : | | | | | | | | | : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 ELGIYVIIDHWILNDGNPNONKEKAKEFFFKEMSLLYGNTPNVTYEIANEPNG-DVNWKRD 179

117 DUGIIVIIDHILSDNDPNYKKEAKDFDEMSELYGDPNVIYEIANEPNGSDVTWDNQ 176

177 IKPYAEVPIVRDNPNIIVGTGTSQDVHHAADQLADPNVMAFHYAGTHQNL 236

QY 237 RQVDYALDQAAIFVSEWGTSAATGGGVFLDEAQVWIDFMDERNLSWANWSLTHKDES 296

Db 300 SSALPKGASKGGRSLDLSASGTFYRNLGTDKTDIDPETSCKDPTQENG 353
 QY 297 SAALMPGANPTGGTAEALSPSGTFVREKIREASIPSPDP-TPPSDGPDPG 349

RESULT 7
 ID GUNW-ERWCA STANDARD; PRT: 504 AA.
 AC Q93395;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE ENDOGLUCANASE VI PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE VI)
 DE (CELLULOSE VI).

OC ERWINIA CARTOVORA.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
 OC ERWINIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SCC3193;
 RX MEDLINE; 95231512.
 RA MAE A., HEIKINHEIMO R., PALVA E.T.;
 RT "Structure and regulation of the Erwinia carotovora subspecies
 RT carotovora SCC3193 cellulase gene celv1 and the role of cellulase in
 RT phytopathogenicity";
 RL MOL. GEN. GENET. 247:17-26(1995).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X79241; G493493; --
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR PFAM; PF00150; cellulase; 1.
 DR PFAM; PF00942; CBD_3; 1.
 DR HSP; Q06851; INEC
 KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
 FT SIGNAL 1 31
 FT CHAIN 32 504
 FT DOMAIN 32 334
 FT CATALYTIC.
 FT LINKER.
 FT DOMAIN 335 352
 FT DOMAIN 353 504
 FT CELLULOSE-BINDING (BY SIMILARITY).
 FT ACT_SITE 168 168
 FT ACT_SITE 256 256
 FT NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 504 AA; 54963 MW; 60AFD878 CRC32;

Query Match 47.3%; Score 1616; DB 1; Length 504;
 Best Local Similarity 65.8%; Pred. No. 2.51e-302; Indels 2; Gaps 2;
 Matches 210; Conservative 55; Mismatches 52;

Db 35 VETHGQLSIENGRVLDEQGRVQLRGISNGLQWGYDVYVKNKMLRDRDGMGIVFRVAM 94
 QY 32 VEERHGLSISNGELVNERGEQVQLKGMSSHGLQWGYQFVNVESKMLRDRDGMGIVFRAM 91
 Db 95 YTAENGIANPLANKVEAAAGLVYIIDWHITLSDNDPNNTYKAQAKIFFAEMAGL 154
 QY 92 YTSGGYIDDPVSKVKETVEAIDLGIYVIDWHILSDNDPNNTYKEEAKDFDEMSEL 151
 Db 155 YGNSPNVIYEIANPENG-VTWNGOIRPYALEVDTITRSKPDNLIIYSGTWSODIHA 213
 QY 152 YGDPNVIYEIANPENGSDVTWNOIKPYAEVPIVNDPNPNIVYGTGTWSODVHA 211
 Db 214 ADNOLPDPTLYALHFYAGTHGQFLRDRIDYAQSRGAALFVSEWGTSDASNGGPFPLPES 273

QY 212 ADNQLADPNVMTAFHYAGTHGQNLRDQVDYALDQGAALFVSEWGTSAATGGVFLDEA 271
 Db 274 QTWIDFLNRRGISWVNSLSDKSETSAALVAGASKSGGWTEQNLTSGKFVREQIRAGAG 333
 QY 272 QVWIDFMDERNLSWANWSLTHDESSAALMPCANPTGGTAEALSPSGTFVREKIREAS 331

Db 334 LSGGDTPTMPTPTNGNG 352
 QY 332 IPPSD-PTPPSDPGEPPDG 349

RESULT 8
 ID GUN3-BACSU STANDARD; PRT: 499 AA.
 AC P23549;
 DT 01-NOV-1991 (REL. 20, CREATED)
 DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 DE (CARBOXYMETHYL-CELLULOSE) (CMCASE) (CELLULOSE).
 OS BACILLUS SUBTILIS.
 OS BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
 OC BACILLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BSE616;
 RX MEDLINE; 91299280.
 RA PARK S.H., KIM H.K., PACK M.Y.;
 RT "Characterization and structure of the cellulase gene of Bacillus
 RT subtilis BSE616";
 RL AGRIC. BIOL. CHEM. 55:441-448(1991).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; D01057; G216388; --
 DR FIR; JN0111; JN0111.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR PFAM; PF00150; cellulase; 1.
 DR PFAM; PF00942; CBD_3; 1.
 DR HSP; Q06851; INEC
 KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
 FT SIGNAL 1 29
 FT CHAIN 30 499
 FT ENDOGLUCANASE.
 FT ACT_SITE 169 169
 FT ACT_SITE 257 257
 FT NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 350 499
 FT CELLULOSE-BINDING (BY SIMILARITY).
 SQ SEQUENCE 499 AA; 55169 MW; 4391DCA5 CRC32;

Query Match 46.7%; Score 1598; DB 1; Length 499;
 Best Local Similarity 61.6%; Pred. No. 1.64e-298;
 Matches 218; Conservative 62; Mismatches 68; Indels 6; Gaps 6;

Db 1 MKRSISIFTCLLITLTLMGGMALSPASAGTKTPVAKNGQLSIGKTQLVNRDGRKAVQK 60
 QY 1 MKKITIFAV-LLMTL-ALFST-GNTTADDYSV-VEEHGQLSISNGELVNERGEQVQLK 56
 Db 61 GTSSEGLQWGYGVYVKNKSLKWLDRDDWGLTVFAAAYTADGGIIDNPVSKNKKAEVAAK 120
 QY 57 GMSHGLQWGYGVYVNESKMLRDRDGMGIVFRAMITSSGGYIDDPVSKVKETVEAAI 116
 Db 121 ELGIYVIDWHILNDGNPNQNKKEAKFEFFKEMSSLYGNTPNVIYEIANPENG-DVNWKR 179

```

QY 117 DLGIYIIDHILSDNDPNYKEAKDFDEMSELYGDPNVIYEIANEPNGSDVTWQ 176
Db 180 IKPAEAEVIRKNDPNIIIVGTGTSQVNDRAADQOLKANDYDALHFFYAGTHGQFL 239
QY 177 IKPAEAEVIRKNDPNIIIVGTGTSQVNDRAADQOLKANDYDALHFFYAGTHGQFL 236
Db 240 RDKANYALSAPFVTEWGTSDASNGGVFLDQSRWLYLDSKDTISVNNWNLSDKQES 299
QY 237 RDQVDYALDQGAALFVSEWGTSAATGCGVFLDEAQQWIDFMDERNLSWANWSLTHKDES 296
Db 300 SALKPGKAGTGWRLSDLSASGTFRVRENILGTDKSTKIDIPPAKDKPTQBN 353
QY 297 SAALMPGANPTGGTAEALSFGTFVREKIREASIPSDP-TTPSDGCEPDG 349

RESULT 9
ID GUN_CLOAB STANDARD; PRT; 448 AA.
AC P15704;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULOSE).
GN EGA.
OS CLOSTRIDIUM ACETOBUTYLICUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-P262;
RX MEDLINE; 88268074.
RA ZAPPE H., JONES W.A., JONES D.T., WOODS D.R.;
RT "Structure of an endo-beta-1,4-glucanase gene from Clostridium
RT acetobutylicum P262 showing homology with endoglucanase genes from
RT Bacillus spp.";
RL APPL. ENVIRON. MICROBIOL. 54:1289-1292(1988).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- THE C-TERMINAL REGION OF C.ACETOBUTYLICUM IS NOT REQUIRED FOR
CC ACTIVITY.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M31311; G144790; -.
CC PIR; A27631; A27631.
CC PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
CC PFAM; PF00150; cellulase; 1.
CC CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
CC SIGNAL 1 34
CC CHAIN 35 448 ENDOGLUCANASE.
CC ACT_SITE 175 175 PROTON DONOR (BY SIMILARITY).
CC ACT_SITE 263 263 NUCLEOPHILE (BY SIMILARITY).
CC SEQUENCE 448 AA; 49366 MW; 5EE996E6 CRC32;

Query Match 43.6%; Score 1491; DB 1; Length 448;
Best Local Similarity 54.9%; Pred. No. 7.41e-276;
Matches 197; Conservative 69; Mismatches 88; Indels 5; Gaps 5;

Db 11 FKKTFFLIAYVMMFTVLGNTYKAEATTSFGQOLKVGQSDNSGKPIQLKGMSSH 70
QY 4 ITTIFAVLL-MTALFSI-GNTTAADDYVVVEHGSLSNCELVNERGEQVQLKGMSSH 61
Db 71 GLOWVFNVDYDKFLDKGKGVNIRAAVYNEGYISNPSQKEKIKKIYQDAIDLNM 130
QY 62 GLOWYGFVNVYESKWLKDDWGIVTFRAMTSSGGYIDDPVSV-REKVKETVEAAILDLGI 120

```

```

Db 131 VYLDHILSDNDPNYKEAKDFDEMSELYGDPNVIYEIANEPNGG-TWANDIKPY 189
QY 121 VYLDHILSDNDPNYKEAKDFDEMSELYGDPNVIYEIANEPNGSDVTWQNIKPY 180
Db 190 ANYIIPAIRADPNIIIVGTGTSQVNDRAADQOLKANDYDALHFFYAGTHGQFL 249
QY 181 AEEVPIVRNDPNIIIVGTGTSQVNDRAADQOLKANDYDALHFFYAGTHGQFL 240
Db 250 NYAMSKGIAFVTEWGTSDASNGGVFLDQSRWLYLDSKDTISVNNWNLSDKQES 309
QY 241 DYALDQGAALFVSEWGTSAATGCGVFLDEAQQWIDFMDERNLSWANWSLTHKDES 300
Db 310 KSGSTTGGWTDSDLTSLFVKKSGSNTTSQTSAPTFSLQSGTGYDAQVITLSSD 368
QY 301 MPGANPTGGTAEALSFGTFVREKIREASIPSDP-TTPSDGCEPDGPTPPSD 358

```

RESULT 10

```

ID GUN_BUTFI STANDARD; PRT; 429 AA.
AC P22541;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE A PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULOSE) (EGA).
GN CELA.
OS BUTYRIVIBRIO FIBRISOLVENS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC BUTYRIVIBRIO.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-50.
RC STRAIN-A46;
RX MEDLINE; 91100957.
RA HAZLEWOOD G.P., DAVIDSON K., LAURIE J.I., ROMANIEC M.P.M.,
RA GILBERT H.J.;
RT "Cloning and sequencing of the cella gene encoding endoglucanase A of
RT Butyrivibrio fibrisolvens strain A46.";
RL J. GEN. MICROBIOL. 136:2089-2097(1990).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M37031; G144155; -.
CC PIR; S29044; S29044.
CC PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
CC PFAM; PF00150; cellulase; 1.
CC CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
CC SIGNAL 1 34
CC CHAIN 35 429 ENDOGLUCANASE A.
CC ACT_SITE 249 249 PROTON DONOR (BY SIMILARITY).
CC ACT_SITE 334 334 NUCLEOPHILE (BY SIMILARITY).
CC SEQUENCE 429 AA; 48858 MW; 0FAFC840 CRC32;

```

Query Match

```

Best Local Similarity 25.9%; Score 884; DB 1; Length 429;
Matches 131; Conservative 65; Mismatches 93; Indels 11; Gaps 8;

```

```

Db 114 GKLAEGSHLVADGHEVLLMGVSTHGINVPEYASAEIKSLRTGTWGNVIRLAMYTD 173
QY 36 GQLSTNGELVNERGEQVQLKGMSSHGLQWYGFVNVYESKWLKDDWGIVTFRAMTSS 95
Db 174 YNGCYVACKENGEKUKIDDAVEAATDNDMYIIDWHTFLDADPNYKADAIQFFGMV 233

```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensee@isb-sib.ch).
CC -----
EMBL; Y00540; G41092; -.
DR PIR; S03767; S03767.
DR DB; IAIW; 06-MAY-98.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PFAM; PF00150; cellulase; 1.
KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL; 3D-STRUCTURE.
FT SIGNAL 1 43
FT CHAIN 44 426 ENDOGLUCANASE Z.
FT DOMAIN 44 332 CATALYTIC.
FT DOMAIN 333 366 LINKER.
FT DOMAIN 367 426 CELLULOSE-BINDING.
FT ACT_SITE 176 176 PROTON DONOR.
FT ACT_SITE 263 263 NUCLEOPHILE (BY SIMILARITY).
FT MUTAGEN 141 141 H->A: LOSS OF ACTIVITY.
FT MUTAGEN 176 176 E->A: LOSS OF ACTIVITY.
FT DISULFID 368 425
FT CONFLICT 293 295
FT CONFLICT 350 364
FT FT
FT FT SNA -> QLTPQ (IN REF. 1).
FT TDTVTDEPTTDDPA -> MTPPLNRPQPOTHRQ (IN
REF. 1).
FT THNEAGSIVVKNLYTANKWTATSVPGSDSSWTVQGSCN
-> LITKQANRSTKAICIPQTGTPIHFRAAIPPGRKLV
AVTN (IN REF. 1).
FT FT
FT FT
FT FT
SQ SEQUENCE 426 AA; 46418 MW; FD9E8988 CRC32;

Query Match 23.9%; Score 816; DB 1; Length 426;
Best Local Similarity 40.8%; Pred. No. 1.08e-134;
Matches 127; Conservative 67; Mismatches 106; Indels 11; Gaps 9;

Db 76 WGGEKFYTDATVASLKDKDWKSIVRAAMGVQESGGYQLDPAGNKAKYERVVVDAAIANDMY 135
| | : : : : : | | | | | | | | | | | | | | | | | | | | | | | |
| | : : : : : | | | | | | | | | | | | | | | | | | | | | | | |
QY 65 WYGQ-FVNYESMKWLRRDWDGIVTFRAAM-VTSGGYIDDPSV-KEKYKETVEAAIDLGIY 121

Db 136 AIIQWHSHSAEN-N-RSEAIRFQEWMARKGNPNVIYEIYNELP--QVSWSNTIKPYA 190
| | : : : : : | | | | | | | | | | | | | | | | | | | | | | | |
| | : : : : : | | | | | | | | | | | | | | | | | | | | | | | |
QY 122 VIIDWHILSDNDPNRYKEEAADFDEMSEALGYDPNVYETIANEPNGSDVTWDSNQIKPYA 181

Db 191 EAVISAIRADPDNLIIIVGPSWSQNDEASRDPIINAKNIATYTLHFYAGTHGESLRNKAR 250
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 182 EEVIPVRINDPNPNIVIGTGWSQDVHHADNQLADPNVNIAFHFYAGTHGQNLROQVD 241

Db 251 QALANGIALFVTEGWTVNADGNGVNOTETDAWYTFMRDNNNISANNALNDKNNEGASTYY 310
| | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 242 YALDQGAIFVSEWGTSAACTGDGGVFLDEAQVWIDFENDERNLSWANNSLTHKDESSAALM 301

Db 311 PSDKNLTESKKVKSIQSHPYKAGSAASATDSTDTTDTTTVDPEPTTDTPTACAN 370
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
| | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
QY 302 PGA-NPT-GGWTEAEI SPSETFVREKIREASIPSPDPTPPSDGPDPGE-PDPTPPSD 358

Db 371 ANVPYPNKVSKD 381
QY 359 PGEPAPWDSNQ 369

RESULT 12
ID GUN4_RUMAL STANDARD; PRG; 312 AA.
AC Q07940;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE IV (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULASE)
DE (EG-IV).
GN EG IV.
OS RUMINOCOCCUS ALBUS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC RUMINOCOCCUS.
RN [1]
```

[illegible]

```

CC  -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC  LINKAGES IN CELLULOSE.
CC  -!- PATHWAY: CELLULOSE DEGRADATION.
CC  -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC  HYDROLASES).
CC  -----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; L01577; G154694; -.
CC  PIR; C42360; C42360.
CC  PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
CC  PFAM; PF00150; cellulase; 1.
CC  PFAM; PF00553; CBD_1; 1.
CC  HSP; P07986; IEXH.
CC  KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
CC  FT SIGNAL 1 36
CC  FT CHAIN 37 466 ENDOGLUCANASE E-5.
CC  FT ACT_SITE 299 329 PROTON DONOR (BY SIMILARITY).
CC  FT ACT_SITE 391 391 NUCLEOPHILE (BY SIMILARITY).
CC  SQ SEQUENCE 466 AA; 49760 MW; 114A0A0E CRC32;
CC
CC  Query Match 23.5%; Score 805; DB 1; Length 466;
CC  Best Local Similarity 39.9%; Pred. No. 1.96e-132;
CC  Matches 122; Conservative 81; Mismatches 88; Indels 15; Gaps 11;
CC
Db 165 VERYGKQVCGTQLCDHGPNVLGRMGSTHGIFOWFDCHTDSSLDALAYDWKADIIRLSM 224
QY 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 32 VEHGQLSISNGELNVEGGVQLKGSSHGLQWYGQFVNYESKMWLRDDWGITVFFAAM 91
Db 225 YIQEDGYETNPRGTFRIDOLIDMATARGLYVIVDWHILTPGDPH-YNLDRAKTFFFAEIA 283
QY 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 92 YTSGGYIDDP-SVKEKVKETVEAIDGLIYVIDWHILSDNDPNYK-EKADFFDEMS 149
Db 284 QRHASKTNLYEIANENG--VSWAS-IKSYAEVPIVIRQDPDSVIIVGTGWSLGV 340
QY 150 ELYGDYFNVIYEIANENPGSDVTWNOIKPYAEVPIVIRDNDPNNIIIVGTGTWQ-DV 208
Db 341 SEGSGPAEIAANPNVNASINWAFHYAASHRDNYLNALREASEL-FPVFVTEFTETYTG 399
QY 209 HHAAD-NQLA-DP---NVMYAFHYAGTQGNLRDQVDYALDQGAALFVSEWGTSAATG 262
Db 400 DGANDFQADRYIDLMAERKIGTWKNYSDDFRSGAVQFPQTCASGQWGSGLKASQW 459
QY 263 DGGVFLDEAQWIDFMDERNLSWANWSLTHKDESSAALMFGANPTGG-WTEAELSPSGTF 321
Db 460 VRSLQ 465
QY 11 : : :
QY 322 VREKIR 327
CC
RESULT 14
ID GUN_BACS1 STANDARD; PRT; 800 AA.
AC P06564;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (ALKALINE CELLULOSE).
OS BACILLUS SP. (STRAIN 1139).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87085443.
RA FUKUMORI F., KUDO T., NARAHASHI Y., HORIKOSHI K.;
RT "Molecular cloning and nucleotide sequence of the alkaline cellulase
RT gene from the alkalophilic bacillus sp. strain 1139."

```

RL J. GEN. MICROBIOL. 132:2329-2335(1986).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- ALKALOPHILIC BACILLUS SP., STRAIN 1139, PRODUCES ONE ALKALINE
CC CELLULASE (PH OPTIMUM 9.0). THIS STRAIN IS NOT A TRUE CELLULOXYLIC
CC MICRO-ORGANISM BECAUSE THE ENZYME IS UNABLE TO HYDROLYSE NATIVE
CC CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M15743; G142667; -
CC EMBL: D00066; D1000485; -
CC FIR: A29003; A29003.
CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC PFAM: PF00150; cellulase; 1.
CC KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
CC SIGNAL 1 30
CC CHAIN 31 800
CC FT ACT_SITE 190 190 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 305 305 NUCLEOPHILE (BY SIMILARITY).
CC SQ SEQUENCE 800 AA; 547179BA CRC32;

Query Match 21.6%; Score 739; DB 1; Length 800;
Best Local Similarity 40.5%; Pred. No. 6.52e-119;
Matches 126; Conservative 76; Mismatches 79; Indels 30; Gaps 15;

Db 68 LVDOHGKQLQGMSTHGLQWFFPEILNAYKALANDWESMIRLAMYGENGYASNPGL 127
Qy 45 LVNREGQVQLKGMSSHGLQWYGFVNYESMKWLRDMDGTVFRAAMYSGGYIDDPFS- 103
Db 128 IKSRIKIGDLAIENDMYIVDWHVHAPGDPDPVY-AGAEDFFRDIATLALPNPHIYE 186
Qy 104 VKEKYKETEVAALDGLIYVIDWHILSDNDP-N-IYKEAKDFDEMSELYGDPNVIYE 161
Db 187 LANEPSSNNGGAGIPNNEGNNAVKEYADPIVEMLRDSDNADNIIIVGSPNWSQRPDL 246
Qy 162 IANEP-----NG-SDVTW--D--NQIKPYAEVPIVRDN-DPN-NIVIGTGTWSQDVH 210
Db 247 ADNPIDDDHTHTYVHYFGSHAASTESYPPETPNSERGVNMNTRYALENGVAVFATEW 306
Qy 211 AADNQLADPNVMYAFHYAGTHG-----Q--NL-RDQV--D--YALDQGAIFVSEW 255
Db 307 GTSQANGDGGVDFEADVWIERLNNISWANSWLNKNEVSGAFTPFELGKSNATSLDP 366
Qy 256 GTSATGDDGVDFDEAQVWIDFMDERNLSWANSWLNHKSDESSAALMPGANPTGWTAEAL 315
Db 367 GPDQVWVPEEL 377
Qy 316 SPSSGTFVREKI 326

RESULT 15
ID GUNA_STRLI STANDARD; PRT; 459 AA.
AC P27035;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE CELA PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULOSE).
GN CELA.
OS STREPTOMYCES LIVIDANS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; STREPTOMYCINEAE; STREPTOMYCETACEAE; STREPTOMYCES.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-43.

RC STRAIN=66;
RX MEDLINE: 92246492.
RA THEBERGE M., LACAZE P., SHARECK F., MOROSOLI R., KLUEPFEL D.;
RT "Purification and characterization of an endoglucanase from
RT Streptomyces lividans 66 and DNA sequence of the gene.";
RL APPL. ENVIRON. MICROBIOL. 58:815-820(1992).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- PFM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
CC SITE FOR PROTEOLYSIS.
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M82807; -; NOT_ANNOTATED_CDS.
CC PROSITE: PS00561; CBD_BACTERIAL; 1.
CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC PFAM: PF00150; cellulase; 1.
CC PFAM: PF00553; CBD_1; 1.
CC HSP: P07986; IEXH;
CC KW CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
CC SIGNAL 1 27
CC CHAIN 28 459 ENDOGLUCANASE CELA.
CC FT DOMAIN 136 135 CELLULOSE-BINDING.
CC FT DOMAIN 147 147 LINKER ("HINGE") (PRO-THR BOX).
CC FT DOMAIN 148 357 CATALYTIC.
CC FT ACT_SITE 286 286 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
CC FT DISULFID 31 131 BY SIMILARITY.
CC SQ SEQUENCE 459 AA; 48663 MW; 4035BEC4 CRC32;

Query Match 21.3%; Score 727; DB 1; Length 459;
Best Local Similarity 39.7%; Pred. No. 1.84e-116;
Matches 124; Conservative 71; Mismatches 98; Indels 19; Gaps 15;

Db 156 NGQLHVCVHLCNOYDRPIQLRGMSTHGLQWFGPCYGD-ASLDRLAODWKSDDLRLVAMV 214
Qy 35 HGQLSINGELVNERGEQVQLKGMSSHGLQWYGO-FVNYESMKWLRDMDGTVFRAAMYT 93
Db 215 QEDGYETDPAGFTTSRVNGLVMDAEDRGMYAVIDFHTLTPGDPN-YNLDRATFTSSVAAR 273
Qy 94 SSGGYIDDPVS-KEKVETVEAATDLGIYVIDWHILSDNDPNYK-EAKDFDEMSEL 151
Db 274 -NDKKNIYEIANEPNG--VSWTA-VKSYAEQVPIVIRAADPADVIVYVGTGWSGLGVS 329
Qy 152 YGDPNVIYEIANEPNGSDVTWQIKPYAEVPIVRDNDPNVIVGTGWSQ-DVHH 210
Db 330.GANSEVVPVNTATNIMYAFHFAASHKDYRAVRPAATR-LPLFVSEFGTVSATAMS 388
Qy 211 AA-DNQLAD-P-----NVMYAFHYAGTHGQNLRDQVYALDQGAIFAIFSEWGTSAATGDG 264
Db 389 -VDRSSSVAVMLDLQLKISYANTYSDADGSAAFRGTCGTGTDYSSSGVLTSSGALVK 447
Qy 265 GVFLDEAQVWIDFMDERNLSWANSWLNHKSDESSAALMPGANPTGWTAE-LSPSGTFV 323
Db 448 SRISTTDDFFTS 459
Qy 324 EKRESASIPPS 335

Search completed: Fri Jun 4 09:56:15 1999
Job time : 25 secs.

WQESRLH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 4 09:56:32 1999; MasPar time 29.15 Seconds
874.368 Million cell updates/sec
Tabular output not generated.

Title: >US-08-945-574-1
Description: (1-467) from US08945574.pep
Perfect Score: 3419
Sequence: 1 MKKITTIFAVLLMTLALFSI.....KWTQNPDPGDPGWEPLN 467

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprenb19

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organella
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 49.047; Variance 101.519; scale 0.483

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description	Pred. No.
1	2781	81-3	400	2	085465	0.00e+00
2	2578	75-4	389	2	059232	0.00e+00
3	1649	48-2	499	2	045532	1.33e-297
4	1643	48-1	499	2	052731	1.33e-297
5	1637	47-9	486	2	045430	2.23e-296
6	1622	47-4	501	2	083012	3.72e-295
7	1594	46-6	387	2	031029	4.24e-292
8	942	27-6	570	2	059665	2.14e-286
9	895	26-2	749	2	059154	9.07e-155
10	850	24-9	557	2	094622	2.07e-145
11	853	24-9	930	2	059290	1.78e-136
12	840	24-6	481	2	066084	4.52e-137
13	830	24-3	635	2	066065	1.71e-134
14	749	21-9	783	2	045554	1.63e-132
15	737	21-6	476	5	018453	1.58e-116
16	739	21-6	821	2	059241	3.61e-114
17	707	20-7	319	5	077449	1.46e-114
18	701	20-5	391	5	077094	2.78e-108
19	701	20-5	392	5	044078	4.16e-107
20	697	20-4	319	5	061595	2.52e-106

21	699	20-4	470	5	016028	BETA-1,4-ENDOGLUCANASE	1.02e-106
22	691	20-2	319	5	018454	BETA-1,4-ENDOGLUCANASE	3.77e-105
23	675	19-7	494	2	086099	CELLULOSE PRECURSOR (E	5.05e-102
24	649	19-0	148	2	Q03882	ENDOGLUCANASE PRECURSO	5.92e-97
25	563	16-5	621	2	Q07653	CELLULOSE B.	2.64e-80
26	507	14-8	910	3	P87211	ENDOGLUCANASE PRECURSO	1.39e-69
27	202	5-9	363	2	Q07652	CELLULOSE A.	1.37e-14
28	193	5-6	516	2	Q06185	MANNANASE.	3.83e-13
29	176	5-1	357	2	Q59441	CELODEXTRINASE PRECUR	1.80e-10
30	170	5-0	865	2	Q43919	CHITINASE A PRECURSOR	1.50e-09
31	169	4-9	287	2	Q59144	CHITINASE PRECURSOR (E	2.13e-09
32	163	4-8	507	2	Q50506	PROBABLE HYDROLASE.	1.73e-08
33	160	4-7	252	2	Q06614	PUTATIVE SECRETED CHIT	4.87e-08
34	160	4-7	474	2	Q59142	CHITINASE PRECURSOR (E	4.87e-08
35	161	4-7	542	2	Q59145	CHITINASE II PRECURSOR	3.45e-08
36	157	4-6	538	2	Q59143	CHITINASE PRECURSOR (E	1.36e-07
37	154	4-5	729	2	P96168	CHITINASE A.	3.77e-07
38	153	4-5	831	2	Q50076	CHITINASE B (EC 3.2.1.	5.28e-07
39	151	4-4	535	2	Q59141	CHITINASE PRECURSOR (E	1.04e-06
40	151	4-4	665	2	Q48373	CHITINASE PRECURSOR.	1.04e-06
41	141	4-1	288	2	Q52246	ORFC.	2.83e-05
42	139	4-1	294	2	Q50152	CHITINASE C.	5.41e-05
43	138	4-0	268	2	Q54942	ORF IOTA.	7.47e-05
44	137	4-0	695	2	Q52863	EXOCHITINASE.	1.03e-04
45	134	3-9	430	2	Q50590	CHITINASE PRECURSOR (E	2.69e-04

ALIGNMENTS

RESULT 1
ID 085465 PRELIMINARY; PRT; 400 AA.
AC 085465;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ALKALINE CELLULOSE CEL5A (EC 3.2.1.4).
GN CEL5A.
OS BACILLUS AGARADHAERENS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-DSM 8721;
RC BJORNVDAM M.E.;
RL SUBMITTED (MAY-1998) TO ENBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF067428; G3193120; -;
KW HYDROLASE; GLYCOSIDASE.
SQ SEQUENCE 400 AA; 44702 MW; 85AEALF6 CRC32;

Query Match 81.3%; Score 2781; DB 2; Length 400;
Res. Local Similarity 95.6%; Pred. No. 0.00e+00;
Matches 387; Conservative 8; Mismatches 5; Indels 5; Gaps 1;

Db	1	MKKITTIFVLLMTLALFSI	TAADNDVSVEEHGQLSISNGELVNERGEQVQLKGMSS	60
QY	1	MKKITTIFVLLMTLALFSI	TAADNDVSVEEHGQLSISNGELVNERGEQVQLKGMSS	60
Db	61	HGLWYGFVYVESKMKLR	DDWGINVFRAAMYTSSGGYIDDPVSVEKVEKVEAAIDLDI	120
QY	61	HGLWYGFVYVESKMKLR	DDWGINVFRAAMYTSSGGYIDDPVSVEKVEKVEAAIDLDI	120
Db	121	YVIDWHLSDNDPNVYKEE	KDFDEMSLYGDPVNIYEIANEPNGSDVTWGNQIKPY	180
QY	121	YVIDWHLSDNDPNVYKEE	KDFDEMSLYGDPVNIYEIANEPNGSDVTWGNQIKPY	180
Db	181	AEVPIIRNDPNVYKEE	KDFDEMSLYGDPVNIYEIANEPNGSDVTWGNQIKPY	240
QY	181	AEVPIIRNDPNVYKEE	KDFDEMSLYGDPVNIYEIANEPNGSDVTWGNQIKPY	240
Db	241	DYALDQGAIFVSEWGTSA	ATGDGCVFLDEAQVWIDFMDERNLSWANNSLTHKDESSAAL	300
QY	241	DYALDQGAIFVSEWGTSA	ATGDGCVFLDEAQVWIDFMDERNLSWANNSLTHKDESSAAL	300

SEQUENCE FROM N.A.
 STRAIN=79-23;
 YOON K.-H., JUNG K.H.;
 SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 LINKAGES IN CELLULOSE.
 EMBL; AF045482; G2854064; -.
 HYDROLASE; GLYCOSIDASE.
 SEQUENCE 499 AA; 55193 MW; 11418788 CRC32;

Db	61	G	ISSHGLQWGEYVYVNDKSLKWLRRDDNGIIVFRAAMYTADGGYIDNPSVKNVKEAVEAAK	120			
QY	57	G	MSHGLQWYQGVNTYESKWLRRDDNGIIVFRAAMYTSSGGYIDDPSPVKEVKYETVEAAI	116			
Db	121	E	LGIYVIIDWHILNDGNPNONKEKAEFFKEMSSLYGNTPNVIYETANEPNG-DYNNKRD	179			
QY	117	D	LGIYVIIDWHILNDGNPNIIYEEAKDFDEMSLYGDPNVIYETANEPNGSDVTWQNG	176			
Db	180	I	KPYAEVIVIRKNDPDNIIIVGTGTSQDVNDAAADDQDKDANVMYALHFYAGTHGQSL	239			
QY	177	I	KPYAEVIVIRKNDPDNIIIVGTGTSQDVNDAAADDQDKDANVMYALHFYAGTHGQSL	236			
Db	240	R	KANYALSKGAPFVTEWGTSDASONGGVLDQSRWLKYLDKSTISVWNWNLSDKQES	299			
QY	237	R	QVDYALDQGAALFYSEWGTSAATGGGVLDQSRWLKYLDKSTISVWNWNLSDKQES	296			
Db	300	S	SALKPGASTGGWRLSDLSASCTFVRENILGTDKTDIPETPSKDKXTQENG	353			
QY	297	S	SAALMPCANPTGGWTEAELSPSGTFVRENILGTDKTDIPETPSKDKXTQENG	349			
RESULT 6 PRELIMINARY; PRT; 501 AA.							
ID	083012						
AC	083012						
DT	01-NOV-1998	(TREMBREL. 08, CREATED)					
DT	01-NOV-1998	(TREMBREL. 08, LAST SEQUENCE UPDATE)					
DE	01-NOV-1998	(TREMBREL. 08, LAST ANNOTATION UPDATE)					
OS	CELLULOSE						
OS	BACILLUS SP.						
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;						
OC	BACILLUS.						
RC	[1]						
RC	SEQUENCE FROM N.A.						
RC	STRAIN=5H;						
RA	XIANONGNUCH C., OOI T., KINOSHITA S.;						
RT	"Cloning and nucleotide sequence of beta-mannanase and cellulase gene						
RT	from Bacillus sp. 5H."						
RL	SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.						
RL	EMBL: AB016164; D1032673; --						
SQ	SEQUENCE 501 AA; 55422 MW; 5D28F2BD CRC32;						
Query Match							
Best Local Similarity 61.48; Score 1622; DB 2; Length 501;							
Matches 219; Conservative 62; Mismatches 67; Indels 6; Gaps							
Db	1	M	KROVSLIIVCLISILATSLGNGPNVASATGAKTPVAINGQLSIKGTQLINQNGKAVQIK	60			
QY	1	M	KKITTIFAY-LLMT-LALFSGINTTA-ADDYSV-VEEHGQLSISNGELVNERGEQVQLK	56			
Db	61	G	ISSHGLQWYQGVNTYESKWLRRDDNGIIVFRAAMYTADGGYIDNPSVKNVKEAVEAAK	120			
QY	57	G	MSHGLQWYQGVNTYESKWLRRDDNGIIVFRAAMYTSSGGYIDDPSPVKEVKYETVEAAI	116			
Db	121	E	LGIYVIIDWHILNDGNPNONKEKAEFFKEMSSLYGNTPNVIYETANEPNG-DYNNKRD	179			
QY	117	D	LGIYVIIDWHILNDGNPNIIYEEAKDFDEMSLYGDPNVIYETANEPNGSDVTWQNG	176			
Db	180	I	KPYAEVIVIRKNDPDNIIIVGTGTSQDVNDAAADDQDKDANVMYALHFYAGTHGQSL	239			
QY	177	I	KPYAEVIVIRKNDPDNIIIVGTGTSQDVNDAAADDQDKDANVMYALHFYAGTHGQSL	236			
Db	240	R	KANYALSKGAPFVTEWGTSDASONGGVLDQSRWLKYLDKSTISVWNWNLSDKQES	299			
QY	237	R	QVDYALDQGAALFYSEWGTSAATGGGVLDQSRWLKYLDKSTISVWNWNLSDKQES	296			
Db	300	S	SALKPGASTGGWRLSDLSASCTFVRENILGTDKTDIPETPSKDKXTQENG	353			
QY	297	S	SAALMPCANPTGGWTEAELSPSGTFVRENILGTDKTDIPETPSKDKXTQENG	349			
RESULT 7 PRELIMINARY; PRT; 387 AA.							
ID	031029						
AC	031029						

```

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE BETA(1,4)-GLUCAN GLUCANOHYDROLASE PRECURSOR.
GN CELA.

OS ERWINIA CAROTOVORA SUBSP. CAROTOVORA.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ERWINIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LY34;
RA PARK Y.W., LIM S.T., YUN H.D.;
RL MOL. CELLS 0-0(1997).
DR EMBL; AF025768; G2570835; -.
DR PFAM; PF00150; cellulase; 1.
KW SIGNAL; HYDROLASE.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 387 BETA(1,4)-GLUCAN GLUCANOHYDROLASE.
SQ SEQUENCE 387 AA; 42003 MW; 01AD9FB3 CRC32;

Query Match 46.6%; Score 1594; DB 2; Length 387;
Best Local Similarity 60.1%; Pred. No. 2.14e-286;
Matches 211; Conservative 70; Mismatches 67; Indels 3; Gaps 3;

Db 3 MRRQIVKRLTGVVTVTLGMSLSPSALSATPVETHGSLSIENGRLVDQGRKRVOLRGVS 62
QY 1 MKKTTTFVALL-MTALFSGTNTAADDYSVVEHGLSISNGELVNERGEQVQLKMS 59

Db 63 SHGLQWFGDYKKDSMKWLPDWDGINSRVAMYTAAADGYISKPISKVYKAAQAQSLG 122
QY 60 SHGLQWYGFVNESMKWLRDWDGITVFRAAMYTSSGGYIDDPVSKEKVKETVEAAIDIG 119

Db 123 VYIIDWHILSDNPNYKEAKTFEAMAGLYGNSPTVYIEANEPNGG-VTWGGETRP 181
QY 120 IYVIDWHILSDNPNYKEAKDFDEMSELYGDPYNYIEANEPNGSDVTWQIKP 179

Db 182 YALEVTEPIRSKDPNLIIVGTGWSQDTHDAADNQLDPNTLYALHFVAGTHGFLIR 241
QY 180 YAEVPIVRDNPNNIIVGTGWSQDTHDAADNQLDPNNYAFHFYAGTHGNLRDQ 239

Db 242 IDYAQSGRAAIFVSWGTSDASNGSGPPELPSHTWIDFLNRRGSGRVNWSLTKDESSAA 301
QY 240 VDYALDQGAALFVSEMGTSAAATGCGVDFDEAVQVWIDFMDERNLSWANWSLTHKDESSAA 299

Db 302 LAPGASKCGWTEQNLSAGKVFVRAQIRAAATLSGGDPTPTTEPTNPGSG 352
QY 300 LMPGANPTGGWTEAELSPGTFVREKIREASIPPSD-PTPPSDPGPDGP 349

RESULT 8
ID Q59665 PRELIMINARY; PRT; 570 AA.
AC Q59665;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULASE) (ENDOGLUCANASE)
DE (CARBOXYMETHYL CELLULASE).
GN CELE.
OS PSEUDOMONAS FLUORESCENS.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC PSEUDOMONAS.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE; 95366948.
RA HALL J., BLACK G.W., FERREIRA L.M.A., MILLWARD-SADLER S.J.,
RA ALI B.R.S., HAZLEWOOD G.P., GILBERT H.J.;
RT "The non-catalytic cellulose-binding domain of a novel cellulase from
RT Pseudomonas fluorescens subsp. cellulosa is important for the
RT efficient hydrolysis of Avicel."
RL BIOCHEM. J. 309:749-756(1995).
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
DR EMBL; X86798; G806574; -.

DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF00553; CBD_1; 1.
KW HYDROLASE; GLYCOSIDASE.
SQ SEQUENCE 570 AA; 59925 MW; 4B4FA1E4 CRC32;

Query Match 27.6%; Score 942; DB 2; Length 570;
Best Local Similarity 41.1%; Pred. No. 9.07e-155;
Matches 137; Conservative 76; Mismatches 108; Indels 12; Gaps 9;

Db 40 DVAPLSV-QGNKILANGPASFSGMSLFWNTWEGEKKYNAQVYVSWLKSDNNAKLVRRA 98
QY 34 EHGQLSISNGELVNERGEQVQLKMS--SHGLQWYGFVNTYESMKWLRDWDGITVFRAA 90

Db 99 MGVEDEGGYLTDPANKRDRTQVVDAAIANDMVIIIDW--SHNAHQ-YOQAIARFQENA 155
QY 91 M-YTSSGGYIDDPVSKEKVKETVEAAIDIGYIIVDWHILSDNPNYKEAKDFDEMS 149

Db 156 RYKGANNNHYIYIYNEPL--QVSWNTIKPYAQAVIAAIRAIDPONLIIVGTPTWSDQVD 213
QY 150 ELYGDPYNNYIEANEPNGSDVTWQIKPYAEVIVPVRDNDPNIVIVGTGWSQDVH 209

Db 214 VAANDPITGYQNIAYTLHFYAGTHGYURDKAOTALNRGIALFVTEWGSVANGDGAVAN 273
QY 210 HAANDQLAD-PNVYAFHFYAGTHGNLRDQVDYALDQGAALFVSEWGTSAATGCGVEL 268

Db 274 SETNAWVSFMKTNHISNANWALNDKVEGASALVPGASANGWVNSOLTASGALAKSIISG 333
QY 269 DEAQVWIDFMDERNLSWANWSLTHKDESSAALMPGANPTGGWTEAELSPGTFVREKIRE 328

Db 334 WPSYNTSSSSSAVSSOTVSSSQAPVYVSSSS 366
QY 329 SASIPSPDPTP-SDPGPDGPDPPTPPSDPG 360

RESULT 9
ID Q59154 PRELIMINARY; PRT; 749 AA.
AC Q59154;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE (EC 3.2.1.4) (CELLULASE) (ENDO-1,4-BETA-GLUCANASE)
DE (CARBOXYMETHYL CELLULASE).
GN CELD.
OS ANAEROCELLUM THERMOPHILUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC ANAEROCELLUM GROUP; ANAEROCELLUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Z-1320;
RA ZVERLOV V., ASCHERL G., VELIKOVORSKAYA G., BRONNENMEIER K.;
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
DR EMBL; 277855; E257608; -.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF00395; SLH; 3.
KW HYDROLASE; GLYCOSIDASE.
SQ SEQUENCE 749 AA; 85066 MW; C558A863 CRC32;

Query Match 26.2%; Score 895; DB 2; Length 749;
Best Local Similarity 46.5%; Pred. No. 2.07e-145;
Matches 125; Conservative 66; Mismatches 65; Indels 13; Gaps 9;

Db 68 LADQGEIIOLEGMSTHGLQWYGDILNKNFAKLSKDNKDECNVIRLAMYEGEGYASNPIS 127
QY 45 LVNERGEQVQLKMSHGLQWYGFVNTYESMKWLRDWDGITVFRAAMYTSSGGYIDDPVS 104

Db 128 KEKVTEGKLAENDMYIVTDWHLNPGDPAEIKYG-AKDFEKTATSPNDY-HIYE 185
QY 105 KEKVETVEAAIDIGYIIVDWHILSDNPN--IYKEAKDFDEMSELY-GDYPNVIIE 161

Db 186 LCNEPNPNPGVENSIDGKWKYKAYAQPIKRLSILGNONIIIVGSPNWSQRPDFAIQDP 245
QY 186 LCNEPNPNPGVENSIDGKWKYKAYAQPIKRLSILGNONIIIVGSPNWSQRPDFAIQDP 245

```

QY 162 IANEPGSD--V--TWD--NQIKPYAEVPIVRDNDPNNIVGTGTWSQDVHHAADNQ 215
DB 246 INDKNWYVHFVSGTHKVGVEFNKNKAFENGVPIFVSEWGTSLASGDGGPYLDEADK 305
QY 216 LADPNVYAFHFYAGTH--GQN--LRDQDYALDQGAIFVSEWGTSAATGDGGVFLDEAQV 273
DB 306 WLEYLNSYISWVNSLSNKTSAFVP 334
QY 274 WIDFMDERNLSWANSLSLTHKDESSAALMP 302

RESULT 10
ID P94622 PRELIMINARY; PRT; 557 AA.
AC P94622;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENDO-1.4-BETA GLUCANASE ENG (EC 3.2.1.4) (CELLULOSE) (ENDOGLUCANASE)
DE (ENDO-1.4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULOSE).
GN ENG.

OS CLOSTRIDIUM CELLULOVORANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97136706.
RA SHEWITA S.A., ICHI-ISHI A., PARK J.S., LIU C., MALBURG L.M.,
RA DOI R.H.;
RT "Characterization of engF, a gene for a non-cellulosomal Clostridium
cellulovorans endoglucanase.";
RL GENE 182:163-167(1996).
CC -I- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
LINKAGES IN CELLULOSE.
DR EMBL; U37056; G1778709;
DR PFAM; PF00150; cellulase; 1.
KW HYDROLASE; GLYCOSIDASE.
SQ SEQUENCE 557 AA; F4D7C185 CRC32;

Query Match 24.9%; Score 850; DB 2; Length 557;
Best Local Similarity 41.3%; Pred. No. 1.78e-136;
Matches 133; Conservative 76; Mismatches 91; Indels 22; Gaps 13;

DB 63 LCDKDGNIQLRGMSHGLQWFGVNNNAFAALSNDWNSNVIRLAMYAEGGYATNP 122
QY 45 LVNERGEQVQLKGMSSHGLQWGFVNYESMKWLRDDWGITVFRAMYTSSGGYIDDP 104
DB 123 KQTVINGINVAINDMTVIVDWHVLTGPDNPADYK--AMDFEIKSOKYNNPHIYE 180
QY 105 -KEVKETVEAAILDGIYIIVDWHLSNDPN--IYKEAKDFFDEMSELYGDPNVIYE 161
DB 182 CNEPENGCVTNDATGWAQVKSATPIVOLLRNKGNENLIIVGNPFWSPQRPDLAAND 241
QY 163 ANEPGSD--VTWDN----QIKPYAEVPIVRDNDPNNIVGTGTWSQDVHHAADNQ 216
DB 242 NDSNTMYSVHFYPISTVTDNRNAMSNVRYALNHAAGVAFATEWGTSLATGTG 301
QY 217 ADPNVYAFHFYAGTH--GQ--NL--RDQ--V--D--YALDQGAIFVSEWGTSAATGDGGVFL 268
DB 302 AKADWLDFLNGNNSISWNSISNKAALNSLTSLDPSGDKLWADNELLTSGQVYRA 361
QY 269 DEAQVWIDFMDERNLSWANSLSLTHKDESSAAL--MPGANPTGG--WTEAELSPSGTFVRE 324
DB 362 RIKGAYATVPDPT--NQTPAP 382
QY 325 KIRASASIPSPDPTPPSDGPEP 346

RESULT 11
ID Q59290 PRELIMINARY; PRT; 930 AA.
AC Q59290;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)

DE BETA-1,4-ENDOGLUCANASE (EC 3.2.1.4).
GN CELA.
OS CLOSTRIDIUM JOSUI.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RA FUJINO T., FUJINO E., KARITA S., OHMIYA K.;
RT "Revised sequence of cea gene encoding endoglucanase (EG)-1 from
Clostridium josui.";
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; D85526; D1013510;
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF00395; SLH; 1.
KW HYDROLASE; GLYCOSIDASE.
SQ SEQUENCE 930 AA; 101726 MW; B1C65689 CRC32;

Query Match 24.9%; Score 853; DB 2; Length 930;
Best Local Similarity 44.2%; Pred. No. 4.52e-137;
Matches 122; Conservative 71; Mismatches 37; Indels 20; Gaps 14;

DB 62 LCDKDGNIQLRGMSHGLQWFGPEIINNNAFAALSCKDGSNVIRLAMYAEGYSKDP 121
QY 45 LVNERGEQVQLKGMSSHGLQWGFVNYESMKWLRDDWGITVFRAMYTSSGGYIDDP 104
DB 122 IKRVIDIDILAINDMTVIVDWHVLTGPDNPADYK--AMDFEIKSOKYNNPHIYE 180
QY 105 -KEVKETVEAAILDGIYIIVDWHLSNDPN--IYKEAKDFFDEMSELYGDPNVIYE 161
DB 181 LANEPSPNDPGVNDADAGWAK--VKSAYEPIKILRDSNKNLIIVGSPNWSQRPDLAEN 239
QY 162 IANEP--N--G--S--DVT--WDNQIKPYAEVPIVRDNDPNNIVGTGTWSQDVHHAADN 214
DB 240 PINDNTAYSPHFYSGTHKSTSTSDTRGNIMSNARYALEHGVAVFCSEWGTSEAGNNCP 299
QY 215 QLADPNVYAFHFYAGTH--G--QNL--RDQV--D--YALDQGAIFVSEWGTSAATGDGGV 266
DB 300 YLKEADWLEFLNANNISWINNLSLTKNETSGSFIP 335
QY 267 FLDEAQVWIDFMDERNLSWANSLSLTHKDESSAALMP 302

RESULT 12
ID O66064 PRELIMINARY; PRT; 481 AA.
AC O66064;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE.
OS ACTINOMYCES SP. 40.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; ACTINOMYCINAEAE; ACTINOMYCETACEAE; ACTINOMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-40;
RA CHO K.K.;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U94825; G2980982;
SQ SEQUENCE 481 AA; 50751 MW; 6DE4F93F CRC32;

Query Match 24.6%; Score 840; DB 2; Length 481;
Best Local Similarity 37.6%; Pred. No. 1.71e-134;
Matches 120; Conservative 83; Mismatches 107; Indels 9; Gaps 7;

DB 44 GLSAPTILDEHGMPPQLRGASTHGLQWFGQYVYKAAFSRLDRWGLNVMVRLAVYREGY 103
QY 39 SISNGELVNERGEQVQLKGMSSHGLQWGFVNYESMKWLRDDWGITVFRAMYTSSGGY 98
DB 104 LOGSOAQMDKTEIQCVQAATDLGVTIIVDWHVL--NYPNGDATQAESFFKSYAAKYSVG 162
QY 99 IDDPSPV--KEK--VKETVEAAILDGIYIIVDWHLSNDPNVIYKEAKDFFDEMSELYGDP 156
DB 163 NVIFEVCPNEPTPTWYDGSNDIYCYCTMAKIRADAGSDAILLCGTNTWSQDIDAVAGK 222

Qy	157	NVIYEIANEPNGSD--VTWDNQIKPYAEVPIVRONDPNNVIVGTGTSQDVHHAADN	214
Db	223	PLSADGDNIMYUHFHYAAATKKDDLRKAKLTALNAGTFFVFVSFEGLCDASGGNGIQDQSA	282
Qy	215	QL-ADP--NVMYAFHFYAGTHGQNLRDOVDYALDOGAEIFSEWGTSAATGDGVFLDEA	271
Db	283	NAMWTLLAHNNISYAVALNSKAETAAFFKPSVTATSKTGGDLLTPSAIWLNVNTSKRLAD	342
Qy	272	QVWIDFDMERNLWSANSLTHKDESSAALPGANPTGGTWTEAELSPSGTFVREKIRESA	331
Db	343	FADHAAAGSTSGSSKASSG	361
Qy	332	IP-PSDTPPPSDGPDPFG	349
RESULT	13	PRELIMINARY;	PRT; 635 AA.
ID	O66065		
AC	O66065;		
DT	01-AUG-1998 (TREMBUREL, 07, CREATED)		
DT	01-AUG-1998 (TREMBUREL, 07, LAST SEQUENCE UPDATE)		
DT	01-AUG-1998 (TREMBUREL, 07, LAST ANNOTATION UPDATE)		
DE	CWC-XYLANASE (FRAGMENT).		
OS	FIBROBACTER SUCCINOGENES S85.		
OC	BACTERIA; FIBROBACTER GROUP; FIBROBACTER.		
FN	[1]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN-S85;		
RA	CHO K.K.;		
RL	SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.		
KR	EMBL; U94826; G2980984; -.		
DR	XLYAN DEGRADATION; HYDROLASE; GLYCOSIDASE.		
KW	NON_TER	635	
FT	SEQUENCE	635 AA; 65943 MW; 2510AB57 CRC32;	
SQ			
Query Match 24.3%; Score 830; DB 2; Length 635;			
Best Local Similarity 38.0%; Pred. No. 1.63e-132;			
Matches 127; Conservative 80; Mismatches 117; Indels 10; Gaps 8;			
Db	78	GFSAPTVLDEHGQPFOLRGASTGHGVOMFPQYINRDAPQSRLRDEGMVVELALYPREGGY	137
Qy	39	SISINGELNVEGEQVQLKMSSHLGWYGQFVNYESMKWLRRDWGITVFRAAMTSSGGY	98
Db	138	LOGSOAKMDAKIEAVNAANELGMVTLDHVLI-NYNPNGDADKAEFFTRYATKYKNLK	196
Qy	99	IDDFSVK-E-KVETVEAADLGIIYVIDHHILSDNDPNIIYKEARDFFDEMSELCDYP	156
Db	197	NVLVEIDNEPTSTSWYDGSGNDLTYYSKRITKAIATGNOSVICGTNTWSQDVDAVAK	256
Qy	157	NVIYEIANEPNGSD--VTWDNQIKPYAEVPIVRONDPNNVIVGTGTSQDVHHAADN	214
Db	257	PLSADGIGNAYTIHFYAGTHYHDNIKNKLKTALAAGTTPVFVSFEGITDASGWIIDIANA	316
Qy	215	QL-ADP--NVMYAFHFYAGTHGQNLRDOVDYALDOGAEIFSEWGTSAATGDGVFLDEA	271
Db	317	NMWMTLLTRNNISYAVALNSCLNKGBGSFKLESTSKTSKVTGSELSTSGIWLVTSTRIOA	376
Qy	272	QVWIDFDMERNLWSANSLTHKDESSAALPGANPTGGTWTEAELSPSGT-FVREKIRESA	330
Db	377	MVDVSVSGSGTGGET-PTDPTPDPTDPTDPTDPT	409
Qy	331	SIPPSDTPPSDGPDPGPDPTPPSDPGEYPA	364
RESULT	14	PRELIMINARY;	PRT; 783 AA.
ID	Q45554		
AC	Q45554;		
DT	01-NOV-1996 (TREMBUREL, 01, CREATED)		
DT	01-NOV-1996 (TREMBUREL, 01, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998 (TREMBUREL, 08, LAST ANNOTATION UPDATE)		
DE	ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)		
OS	(ALKALINE CELLULASE).		
OS	BACILLUS SP.		

2
2
2
2
-
-
-
2
3
4
2
2
2

[W][O][R][D][S]

(TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 4 09:58:07 1999; MasPar time 20.71 Seconds
903.345 Million cell updates/sec

Tabular output not generated.

Title: >US-08-945-574-1
Description: (1-467) from US08945574.pep
Perfect Score: 3419
Sequence: 1 MKKITTIFAVLLMTLALFSI.....KWTQNEPCDYPGPWEPLN 467

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 49.622; Variance 115.423; scale 0.430

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2828	82.7	409	2	B25156 cellulase (EC 3.2.1.4	0.00e+00
2	2803	82.0	488	2	A25156 cellulase (EC 3.2.1.4	0.00e+00
3	1664	48.7	505	2	S39962 endoglucanase - Erwin	4.29e-267
4	1660	48.6	508	2	A26874 cellulase (EC 3.2.1.4	2.28e-266
5	1649	48.2	499	2	A27198 cellulase (EC 3.2.1.4	2.26e-264
6	1641	48.0	508	2	G69593 cellulase (EC 3.2.1.4	6.37e-263
7	1637	47.9	486	2	I40548 bifunctional cellulases	3.39e-262
8	1616	47.3	504	2	S54744 cellulase (EC 3.2.1.4	2.17e-258
9	1598	46.7	499	2	JN0111 cellulase (EC 3.2.1.4	3.97e-255
10	1491	43.6	448	2	A27631 cellulase (EC 3.2.1.4	9.24e-236
11	942	27.6	570	2	S56132 cellulase (EC 3.2.1.4	2.65e-137
12	884	25.9	429	2	S29044 endoglucanase A precu	4.93e-127
13	850	24.9	557	2	JC3487 cellulase (EC 3.2.1.4	4.96e-121
14	805	23.5	466	2	C42360 cellulase (EC 3.2.1.4	4.16e-113
15	793	23.2	428	2	S03767 cellulase (EC 3.2.1.4	5.33e-111
16	749	21.9	783	2	JC5467 cellulase (EC 3.2.1.4	2.75e-103
17	739	21.6	800	2	A29003 cellulase (EC 3.2.1.4	1.54e-101
18	739	21.6	822	2	JT0611 cellulase (EC 3.2.1.4	1.54e-101
19	690	20.2	357	2	PC4404 cellulase (EC 3.2.1.4	5.42e-93
20	690	20.2	941	2	S29043 cellulase (EC 3.2.1.4	5.42e-93
21	613	17.9	825	2	JS0174 cellulase (EC 3.2.1.4	1.17e-79
22	196	5.7	516	2	JE0134 mannan endo-1,4-beta-	8.75e-12
23	186	5.4	584	2	JQ1229 cellulase (EC 3.2.1.4	2.26e-10

24	179	5.2	900	2	JH0157 cellulase (EC 3.2.1.4	2.13e-09
25	170	5.0	799	2	PC4106 chitinase (EC 3.2.1.1	3.63e-08
26	170	5.0	820	2	A40633 chitinase (EC 3.2.1.1	3.63e-08
27	169	4.9	287	2	S65765 chitinase (EC 3.2.1.1	4.97e-08
28	160	4.7	474	2	S65763 chitinase (EC 3.2.1.1	7.98e-07
29	161	4.7	542	2	I39540 chitinase (EC 3.2.1.1	5.88e-07
30	157	4.6	538	2	S65764 chitinase (EC 3.2.1.1	1.99e-06
31	153	4.5	831	3	T00323 chitinase (EC 3.2.1.1	6.64e-06
32	151	4.4	535	2	S65762 chitinase (EC 3.2.1.1	1.21e-05
33	141	4.1	288	2	JH0204 hypothetical 30.5K pr	2.28e-04
34	139	4.1	826	2	A45559 sporozoite surface pr	4.05e-04
35	138	4.0	268	2	S45091 hypothetical protein	5.39e-04
36	128	3.7	897	2	E65127 hypothetical 97.1 kD	8.94e-03
37	122	3.6	699	2	C43674 US4 protein - human h	4.56e-02
38	124	3.6	1331	2	A48954 mannan endo-1,4-beta-	2.66e-02
39	119	3.5	285	2	A41826 probable pheromone-re	1.01e-01
40	121	3.5	332	1	OZQOMB circumsporozoite prot	5.96e-02
41	115	3.4	348	1	OZQOMB circumsporozoite prot	2.88e-01
42	116	3.4	449	2	A24993 cellulase (EC 3.2.1.4	2.23e-01
43	117	3.4	825	2	JC4163 DNA-binding protein 5	1.72e-01
44	116	3.4	917	2	S09646 collagen alpha 2(VI)	2.23e-01
45	116	3.4	1018	1	CGHU2A collagen alpha 2(VI)	2.23e-01

ALIGNMENTS

RESULT 1

ENTRY B25156 #type complete

TITLE cellulase (EC 3.2.1.4) 2 - Bacillus sp.

ALTERNATE_NAMES endo-1,4-beta-glucanase

ORGANISM #formal_name Bacillus sp.

DATE 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 20-Mar-1998

ACCESSIONS B25156

REFERENCE A91825

#authors Fukumori, F.; Sashihara, N.; T.; Horikoshi, K.

#journal J. Bacteriol. 1986; 168:499-488

#title Nucleotide sequence of two cellulase genes from alkalophilic Bacillus sp. strain N-4 and their strong homology

#cross-references PMID:87056924

#accession B25156

#molecule_type DNA

#residues 1-409 #label FUK

#cross-references GB:M14729; NID:g142655; PID:g142656

#experimental_source strain N-4, plasmid pNK2

FUNCTION

#description Hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose and lichenin; can hydrolyze such linkages in beta-D-glucans that also contain 1,3-linkages

#pathway cellulose degradation

KEYWORDS glycosidase; hydrolase; polysaccharide degradation

SUMMARY #length 409 #molecular-weight 45690 #checksum 9198

Query Match Score 2828; DB 2; Length 409;
Best Local Similarity 94.6%; red. No. 0.00e+00;
Matches 387; Conservative 12; Mismatches 6; Indels 4; Gaps 2;

Db 1 MKKITTIFVLLMTLALFISGNTAAADDYVVVEEHQQLSINGELVNDGEPVOLKGMSS 60
|||||
Qy 1 MKKITTIFVLLMTLALFISGNTAAADDYVVVEEHQQLSINGELVNDGEPVOLKGMSS 60
|||||

Db 61 HGLQWYGFVNTYESMKWLRDDMGITVFRAMYTSSGGYIEDPSVKKEAVEAIDLGI 120
|||||

Qy 61 HGLQWYGFVNTYESMKWLRDDMGITVFRAMYTSSGGYIEDPSVKKEAVEAIDLGI 120
|||||

Db 121 YVIIDWHIISDNDPNYIEKAEKDFDEMSGLYGDYPNVYEIETANEPNGSDVTWNOIKPY 180
|||||

Qy 121 YVIIDWHIISDNDPNYIEKAEKDFDEMSGLYGDYPNVYEIETANEPNGSDVTWNOIKPY 180
|||||

Db 181 AEEVPIVRNDPNNTIIIVGTGTWSODVHHAADNQLTDPNVMFAPHFYAGTHGQNLRDQV 240
|||||

Qy 181 AEEVPIVRNDPNNTIIIVGTGTWSODVHHAADNQLTDPNVMFAPHFYAGTHGQNLRDQV 240
|||||


```

Db 241 DYALDQGAALFVSEWGTSEATGDCGVFLDEAQWIDFMDERNLNSWANWSLTHKDESSAAL 300
QY 241 DYALDQGAALFVSEWGTSEATGDCGVFLDEAQWIDFMDERNLNSWANWSLTHKDESSAAL 300

Db 301 MPGASPTGCGWTEAELSPSGTFFVREKIRASATPPSDPTPPSDPDGCEPDPGEPDPTPP 360
QY 301 MPGANPTGCGWTEAELSPSGTFFVREKIRASATPPSDPTPPSDPDGCEPDPGEPDPTPP 356

Db 361 SDPGDYPAWDPNTIYTDIEIVYHNGOLQWAKWWTQNOEGDPDYGWPWEPLN 409
QY 357 SDPGDYPAWDPNTIYTDIEIVYHNGOLQWAKWWTQNOEGDPDYGWPWEPLK 405

RESULT 2 A25156 #type complete
ENTRY cellulase (EC 3.2.1.4) 1 - Bacillus sp.
TITLE endo-1,4-beta-glucanase
ALTERNATE_NAMES #formal_name Bacillus sp.
ORGANISM 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change
DATE 20-Mar-1998
ACCESSIONS A25156
REFERENCE A91825
#authors Fukumori, F.; Sashihara, N.; Kudo, T.; Horikoshi, K.
#journal J. Bacteriol. (1986) 168:479-485
#title Nucleotide sequences of two cellulase genes from alkalophilic
Bacillus sp. strain N-4 and their strong homology.
#cross-references MUID:87056924
#accession A25156
#molecule_type DNA
#residues 1-488 #label FUK
#cross-references GB:M14781; GB:X53449; NID:gl42659; PID:gl42660
#experimental_source strain N-4, plasmid pNK1
FUNCTION
#description hydrolysis of 1,4-beta-D-glucosidic linkages in
beta-D-glucans such as cellulose and lichenin; can
hydrolyze such linkages in beta-D-glucans that also contain
1,3-linkages
#pathway cellulose degradation
KEYWORDS glycosidase; hydrolase; polysaccharide degradation
SUMMARY #length 488 #molecular-weight 54264 #checksum 2043

Query Match 82.0%; Score 2803; DB 2; Length 488;
Best Local Similarity 84.0%; Pred. No. 0.00e+00;
Matches 414; Conservative 30; Mismatches 18; Indels 31; Gaps 9;

Db 1 MKKLTTFIVFTLAL-LF-VGNSTSANNGSVVEQNGQLSIQNGVLNHEGDPVOLKGMSS 58
QY 1 MKKLTTFIVFTLAL-LF-VGNSTSANNGSVVEQNGQLSIQNGVLNHEGDPVOLKGMSS 60

Db 59 HGLQWYGFVNYDSIKWLDDWGTIVFRAAMYTSGGYIEDPSVKEKVEAYEAAIDLGI 118
QY 61 HGLQWYGFVNYESMKWLDDWGTIVFRAAMYTSGGYIEDPSVKEKVEAYEAAIDLGI 120

119 YVIDDWHLLSDNDPNIIYEAKKEFFDEMSALYGDYPNVIYEIANEPNGHNVWDSHIRPY 178
QY 121 YVIDDWHLLSDNDPNIIYEAKKEFFDEMSALYGDYPNVIYEIANEPNGSDVTDNQIRPY 180

179 AEEVIPVIRANDPNVIVTGVATVSODVHEAADNOLDPNVMYAFHFYAGTHGQQLRNQV 238
QY 181 AEEVIPVIRANDPNVIVTGVATVSODVHEAADNOLDPNVMYAFHFYAGTHGQQLRNQV 240

239 DYALSRGAALFVSEWGTSAATGDCGVFLDEAQWIDFMDERNLNSWANWSLTHKDESSAAL 298
QY 241 DYALDQGAALFVSEWGTSAATGDCGVFLDEAQWIDFMDERNLNSWANWSLTHKDESSAAL 300

299 MPGANPTGCGWTEAELSPSGTFFVREKIRASATPPSDPTPPSDPDGCEPDPGEPDPTPP 355
QY 301 MPGANPTGCGWTEAELSPSGTFFVREKIRASATPPSDPTPPSDPDGCEPDPGEPDPTPP 360

356 EXPADPNQIYTDIEIVYHNGOLQWAKWWTQNOEGCANQYGPWEPLGDPAPPSPSPDP 415
QY 361 EXPADPNQIYTDIEIVYHNGOLQWAKWWTQNOEGC-DPYGPWEPL-K----SDP-D----S 410

```

```

Db 416 EPEPDGEPDGPDPGEPDPTPPSDPGYPAWDPNTIYTDIEIVYHNGOLQWAKWWTQNO 475
QY 411 -----GEPDPTPPSDPGYPAWDSNQIYTDIEIVYHNGOLQWAKWWTQNO 454

Db 476 EPGDYGPWPEPLN 488
QY 455 EPGDYGPWPEPLN 467

RESULT 3 S39962 #type complete
ENTRY endoglucanase - Erwinia carotovora
TITLE #formal_name Erwinia carotovora
ORGANISM 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
DATE 09-Sep-1997
ACCESSIONS S39962
REFERENCE S39962
#authors Cooper, V.J.C.; Salmond, G.P.C.
#journal Mol. Gen. Genet. (1993) 241:341-350
#title Molecular analysis of the major cellulase (CelV) of Erwinia
carotovora: evidence for an evolutionary "mix-and-match" of
enzyme domains.
#accession S39962
#status preliminary
#molecule_type DNA
#residues 1-505 #label COO
#cross-references EMBL:X76000; NID:g434941; PID:g434942
SUMMARY #length 505 #molecular-weight 54899 #checksum 7815

Query Match 48.7%; Score 1664; DB 2; Length 505;
Best Local Similarity 67.7%; Pred. No. 4.29e-267;
Matches 216; Conservative 50; Mismatches 51; Indels 2; Gaps 2;

Db 35 VETHQLSINGRVLDQGRVQLRGISGHQWFGDYVYKDSMKWLDRDWDGIVNFRVAM 94
QY 32 VEEHQLSISNGELYNERGEQVQLKGMSSHGLWQGFVYVYESMKWLDRDWDGIVFRAAM 91

Db 95 YTAADGYISNPLANKVKEAVAAASGLVYIIDWHILSDNDPNIIYKAQAKFFAEMAGL 154
QY 92 YTSGGYIDDPVSKVKETVEAIDLGIYVIDWHILSDNDPNIIYKEEAKDFFDEMSEL 151

Db 155 YCSSPNVIYEIANEPNGG-VTWNGQIRPYALEVYTDITRSKDPDNLIIYGTGWSQIDHA 213
QY 152 YGDYPNVIYEIANEPNGSDVTWDNQIKPYAEVPIVRDNDPNIIYVIGTGTWSQDVHHA 211

Db 214 ADNQLPDPNTYALHFYAGTHGQFLRDIDYAOISGAAIFFVSEWGTSDASGNGGFLPES 273
QY 212 ADNQLADPNMYAFHFYAGTHGQNLRDQVDYALDQGAALFVSEWGTSAATGOGVFLDEA 271

Db 274 QTWIDFLNNGYSWNWSLTDKSEASALAPGASKSGGWTEQNLSITSGKFVREQIRAGAN 333
QY 272 QWIDFMDERNLNSWANWSLTHKDESSAALMPCANPTGGWTEAELSPSGTFFVREKIRAS 331

Db 334 LGGGDTPTTPTPTPNPGNG 352
QY 332 IPPSD-PTPPSDPGEPDPG 349

RESULT 4 A26874 #type complete
ENTRY cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain
TITLE DLG)
ALTERNATE_NAMES endo-1,4-beta-glucanase
ORGANISM #formal_name Bacillus subtilis
DATE 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change
10-Jul-1998
ACCESSIONS A26874; B26874
REFERENCE A26874
#authors Robson, L.M.; Chambliss, G.H.
#journal J. Bacteriol. (1987) 169:2017-2025
#title Endo-beta-1-4-glucanase gene of Bacillus subtilis DLG.
#cross-references MUID:87194581

```

```

#accession A26874
#molecule_type DNA
#residues 1-508 #label ROBI
##cross-references GB:M16185; NID:g143007; PID:g143008
##experimental_source strain DLG
#accession B26874
#molecule_type protein
#residues 39-53 #label ROB2
##experimental_source strain DLG
#note the authors believe Met-1 and Met-2 may be alternate
initiators
COMMENT The low molecular weight of the mature protein suggests
carboxyl-terminal processing as well as removal of the signal
sequence.

FUNCTION
#description hydrolysis of 1,4-beta-D-glucosidic linkages in
beta-D-glucans such as cellulose and lichenin; can
hydrolyze such linkages in beta-D-glucans that also contain
1-3-linkages
#pathway cellulose degradation
KEYWORDS alternative initiators; extracellular protein; glycosidase;
hydrolase; polysaccharide degradation
FEATURE
1-38 #domain (or 2-38) signal sequence #status predicted
#label SIG
SUMMARY #length 508 #molecular-weight 56404 #checksum 5520
Query Match 48.6%; Score 1660; DB 2; Length 508;
Best Local Similarity 62.7%; Pred. No. 2.28e-266;
Matches 222; Conservative 63; Mismatches 63; Indels 6; Gaps 6;

Db 10 MKRSISFTICLLIATVLTGGLQSPASAAAGTKTPVAKNGQLSIKGTQLVNRDKAVOLK 69
QY 1 MKKIITIFAV-LLMT-LALFSI-GNTTAADDYSV-VEEHGQLSISNGELVNERGEQVQLK 56
Db 70 GISSHGLQYGFVYKNSLKLWLRDDGKITVFRAMYTADGGYIDNPSVKNKVEAEAAK 129
QY 57 GMSHGLQYGFVYKNSLKLWLRDDGKITVFRAMYTADGGYIDNPSVKNKVEAEAAK 116
Db 130 ELGIYVIDWHILNDGNPNQKAKDFKEMSSLYGNTPNVIYEIANEPNG-DVNNKRD 188
QY 117 DLGIYVIDWHILNDGNPNQKAKDFKEMSSLYGNTPNVIYEIANEPNGSDVTWQNL 176
Db 189 IKPYAEVISVIRKNDPNIIIVGTGTSQDVDAADQLKDNVYALHFAHFGTHGQSL 248
QY 177 IKPYAEVIVIRNDPNIIIVGTGTSQDVDAADQLKDNVYALHFAHFGTHGQNL 236
Db 249 RKANVALSKGAPFVTEWGTSDASNGVFLDQSRWLNLYLDSKNISWYNNLSKQES 308
QY 237 RQVDYALDQGAIFVSEWGTSAATGGVFLDEAQQWIDFMDERNLSWANWSLTHKDES 296
Db 309 SSALKPGASKTGWPLDITLSTGTFVRENIRGTGDKTDPVETPAQDNPTQKNG 362
QY 297 SAALMPGANPTGGTAEELSPSGTFVREKIRASIPSPDP-TPPSDPGEPDPG 349

RESULT 5
ENTRY A27198 #type complete
TITLE cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain
IFO3034)
ALTERNATE_NAMES endo-1,4-beta-glucanase
ORGANISM #formal_name Bacillus subtilis
DATE 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change
09-Sep-1997
ACCESSIONS A27198
REFERENCE Nakamura, A.; Uozumi, T.; Beppu, T.
#authors Eur. J. Biochem. (1987) 164:317-320
#journal Nucleotide sequence of a cellulase gene of Bacillus subtilis.
#cross-references MUID:87190397
#accession A27198
#molecule_type DNA
#residues 1-499 #label NAK

```

```

##cross-references GB:M28332; NID:g142670; PID:g142671
##experimental_source strain IFO3034
FUNCTION
#description hydrolysis of 1,4-beta-D-glucosidic linkages in
beta-D-glucans such as cellulose and lichenin; can
hydrolyze such linkages in beta-D-glucans that also contain
1-3-linkages
#pathway cellulose degradation
KEYWORDS extracellular protein; glycosidase; hydrolase; polysaccharide
degradation
FEATURE
1-36 #domain signal sequence #status predicted #label SIG
SUMMARY #length 499 #molecular-weight 55075 #checksum 3211
Query Match 48.2%; Score 1649; DB 2; Length 499;
Best Local Similarity 62.7%; Pred. No. 2.26e-264;
Matches 222; Conservative 62; Mismatches 64; Indels 6; Gaps 6;

Db 1 MKRSISFTICLLIATVLTGGLQSPASAAAGTKTPAKNGQLSIKGTQLVNRDKAVOLK 60
QY 1 MKKIITIFAV-LLMT-LALFSI-GNTTAADDYSV-VEEHGQLSISNGELVNERGEQVQLK 56
Db 61 GISSHGLQYGFVYKNSLKLWLRDDGKITVFRAMYTADGGYIDNPSVKNKVEAEAAK 120
QY 57 GMSHGLQYGFVYKNSLKLWLRDDGKITVFRAMYTADGGYIDNPSVKNKVEAEAAK 116
Db 121 ELGIYVIDWHILNDGNPNQKAKDFKEMSSLYGNTPNVIYEIANEPNG-DVNNKRD 179
QY 117 DLGIYVIDWHILNDGNPNQKAKDFKEMSSLYGNTPNVIYEIANEPNGSDVTWQNL 176
Db 180 IKPYAEVISVIRKNDPNIIIVGTGTSQDVDAADQLKDNVYALHFAHFGTHGQSL 239
QY 177 IKPYAEVIVIRNDPNIIIVGTGTSQDVDAADQLKDNVYALHFAHFGTHGQNL 236
Db 240 RKANVALSKGAPFVTEWGTSDASNGVFLDQSRWLNLYLDSKNISWYNNLSKQES 299
QY 237 RQVDYALDQGAIFVSEWGTSAATGGVFLDEAQQWIDFMDERNLSWANWSLTHKDES 296
Db 300 SSALKPGASKTGWPLDITLSTGTFVRENIRGTGDKTDPVETPAQDNPTQKNG 353
QY 297 SAALMPGANPTGGTAEELSPSGTFVREKIRASIPSPDP-TPPSDPGEPDPG 349

RESULT 6
ENTRY G69593 #type complete
TITLE cellulase (EC 3.2.1.4) bg1C precursor - Bacillus subtilis
ALTERNATE_NAMES endo-1,4-beta-glucanase
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
05-Dec-1998
ACCESSIONS G69593; A26114; I40353; S24239; S49103; I39803
REFERENCE A69580
#authors Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Berto, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Dueterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,

```

V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rev, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Takakoshi, A.; Tanaka, T.; Terpestra, P.; Tononi, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weltzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 350:249-256

#title The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

#cross-references MUID:98044033

#accession G69593

#molecule_type DNA

#status nucleic acid sequence not shown; translation not shown

#residues 1-508 #label KUN

#cross-references GB:99113; GB:AL009126; NID:g2634090; PID:e1183471; PID:g2634196

#experimental_source strain 168

REFERENCE A26114

#authors Mackay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moraneli, F.; Seligy, V.

#journal Nucleic Acids Res. (1986) 14:9159-9170

#title Structure of a *Bacillus subtilis* endo-beta-1,4-glucanase gene.

#cross-references MUID:87066783

#accession A26114

#molecule_type DNA

#residues 10-508 #label MAC

#experimental_source strain PAP115

#note part of this sequence, including the amino end of the mature form, was confirmed by peptide sequencing

REFERENCE I40353

#authors Lindahl, V.; Aa, K.; Tronsmo, A.

#journal Antonie Van Leeuwenhoek (1994) 66:327-332

#title Nucleotide sequence of an endo-beta-1,4-glucanase gene from *Bacillus subtilis* CK-2.

#cross-references MUID:95225656

#accession I40353

#status preliminary; translated from GB/EMBL/DBJ

#molecule_type DNA

#residues 10-291, 'N', 293-508 #label LIN1

#cross-references EMBL:X67044; NID:g39776; PID:g39777

REFERENCE S24239

#authors Lindahl, V.; Aa, K.

#submission submitted to the EMBL Data Library, June 1992

#accession S24239

#molecule_type DNA

#residues 10-291, 'N', 293-508 #label LIN2

#cross-references EMBL:X67044; NID:g39776; PID:g39777

#experimental_source strain CK-2

REFERENCE S49103

#authors Wolf, M.; Geczi, A.; Borris, R.

#submission submitted to the EMBL Data Library, December 1993

#description Genes encoding beta glucan-hydrolyzing enzymes in *Bacillus subtilis*: construction of strains deficient in lichenase and cellulase activity.

#accession S49103

#molecule_type DNA

#residues 10-508 #label WOI

#cross-references EMBL:Z39076; NID:g509266; PID:g509267

REFERENCE I39803

#authors Seo, Y.S.; Lee, Y.H.; Pek, U.H.; Kang, H.

#journal Korean J. Microbiol. (1986) 24:236-242

#title Analysis on the nucleotide sequence of the signal region of *Bacillus subtilis* extracellular cellulase gene.

#accession I39803

#status preliminary; translated from GB/EMBL/DBJ

#molecule_type DNA

#residues 10-13, 'V', 15, 'N', 17-18, 'V', 20-21, 'F', 23, 'A', 25-26, 'AI', 29-31, 'P', 33, 'PQ', 36-38, 'V', 40, 'E', 42-45, 'SI', 48-58, 'I', 60, 'Q', 62-80, 'DF', 83-86, 'T', 88-103, 'I', 105-157 #label SEO

#cross-references GB:M38634; NID:g142657; PID:g142658

#experimental_source strain ATCC 6633

COMMENT The low molecular weight of the mature protein suggests carboxyl-terminal processing as well as removal of the signal sequence.

GENETICS bglc

FUNCTION hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose and lichenin; can hydrolyze such linkages in beta-D-glucans that also contain 1,3-linkages

#pathway cellulose degradation

KEYWORDS extracellular protein; glycosidase; hydrolase; polysaccharide degradation

FEATURE 1-38

39-508

SUMMARY #domain signal sequence #status predicted #label SIG\ #product cellulase #status predicted #label MAT #length 508 #molecular-weight 56504 #checksum 4569

Query Match 48.0%; Score 1641; DB 2; Length 508; Best Local Similarity 62.7%; Pred. No. 6.37e-263; Matches 222; Conservative 60; Mismatches 56; Indels 6; Gaps 6;

Db 10 MKRSISFTICLLITLTGGMIASTPASAGTKTPVAKNGQLSKGTOLVNRDQKAVOLK 69

QY 1 MKKITIFAV-LLMTL-ALFS-IGNTAADDSV-VEEHGQLSINGELNVEGEVOLK 56

Db 70 GISSHGLQWYGEYVKNKSLKWLDDMGITVFRAMTADGGYIDNPVKNKVEAEAAK 129

QY 57 GMSHGLQWYGEYVKNKSLKWLDDMGITVFRAMTADGGYIDNPVKNKVEAEAAK 116

Db 130 ELGIYVIDWHLNDGNPNONKEKKEFKEMSLGNTPNVIYEIANEPNG-DVNWKR 188

QY 117 DLGIYVIDWHLNDGNPNONKEKKEFKEMSLGNTPNVIYEIANEPNGDVTW 176

Db 189 IKPYAEVIVIRKNDPNIIIVGTGTSQDVNDADDDQLKDANVMYALHFYAGTHG 248

QY 177 IKPYAEVIVIRKNDPNIIIVGTGTSQDVNDADDDQLKDANVMYALHFYAGTHG 236

Db 249 RDKANALSKGAPFVTEWGTSDASNGGVFLDQSRWLKYLDSKTSISWNWLSKQES 308

QY 237 RDQVDYALDQGAIFVSEWGTSAATGDDGVFLDEAQWIDFMDERNLISWANSLTHKDES 296

Db 309 SSALKPGASKTGWRSLDLSASGTFVRENILGKDTKIDIPETPSKDKPTQENG 362

QY 297 SAALMPCANPTGCTEALSPSGTFVREKRESASIPPSDP-TPPSDPGPDPG 349

RESULT 7

ENTRY I40548 #type complete

TITLE bifunctional cellulase precursor - *Bacillus* sp.

ORGANISM #formal_name *Bacillus* sp.

DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 17-Mar-1999

ACCESSIONS I40548

REFERENCE I40548

#authors Han, S.J.; Yoo, Y.J.; Kang, H.S.

#journal J. Biol. Chem. (1995) 270:26012-26019

#title Characterization of a bifunctional cellulase and its structural gene: the cel gene of *Bacillus* sp. D04 has exo- and endoglucanase activity.

#cross-references MUID:96029707

#accession I40548

#status preliminary; translated from GB/EMBL/DBJ

#molecule_type DNA

#residues 1-486 #label RES

DB 33 VETHGQLS IENGRLVDEQGRKRVQLRG TSSNGLQWVGDI VNRKDSMKWLRDDWG I N V F R V A M 94

Qy

```

ENTRY          S29044      #type complete
TITLE          endoglucanase A precursor - Butyrivibrio fibrisolvens
ORGANISM       #formal_name Butyrivibrio fibrisolvens
DATE          13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
ACCESSIONS    S29044
REFERENCE      S29044
#authors      Hazlewood, G.P.; Davidson, K.; Laurie, J.I.; Romaniec,
               M.P.M.; Gilbert, H.J.
#journal      J. Gen. Microbiol. (1990) 136:2089-2097
#title        Cloning and sequencing of the cels gene encoding
               endoglucanase A of Butyrivibrio fibrisolvens strain A46.
#accession    S29044
#status       Preliminary
#molecule_type DNA
##residues    1-429 #label HAZ
##cross-references EMBL:M37031; NID:g144154; PID:g144155
SUMMARY       #length 429 #molecular-weight 48858 #checksum 682

Query Match    25.9%; Score 884; DB 2; Length 429;
Best Local Similarity 43.7%; Pred. No. 4.93e-127;
Matches 131; Conservative 65; Mismatches 93; Indels 11; Gaps 8;

Db 114 GKLAVERGLVDADGHHVLLMGVSTGHINMYPEVASAETIKSLRDTGWINVIRLAMYTSQ 173
QY 36 QQLSISNGELVNERGEQVQLKGMSSHLQVGFVNTESMKWLDRDDGKITVFRAMYTSS 95

Db 174 YNGCYVAGKENQEKLDIIDDAVEAARDNDMVIIDWHITLMDADPNKYKADAIQFFGEMV 233
QY 96 -GGY-I--DDPS-VKEKVEITVEAIDLGIYVIDWHILSDNDPNYKKEAKDFFDEMS 149

Db 234 RYKNDENVYIEICNPNP--DTTWN--VRRYANEVPIVRNVDA--IILVGTPKWATDLD 289
QY 150 ELYGDPNVIYEIANEPNGSDVTWQNIKPYAEVPIVRNDPNNVIVGTGTSQDVH 209

Db 290 SVLDKPLDFNIMYTFYFAGTHHKAERNALRDALDGLPFISEYGLVDADGDNLEK 349
QY 210 HAADNQLADPNVAFHYFAGTHGQNLQDQVYALDQGAALFVSEWGTSAATGGGVFLD 269

Db 350 EADYWDIMIRKEYGVSSCMNLSNKGSAAMINADCKLSDFTEEDLSAMWLIDQISQ 409
QY 270 EAQWIDFM--DERNLWANSLTHKDESSAALMGANPTGGTAEIPLSPSTFYREKIRE 328

RESULT 13
ENTRY      JC5487      #type complete
TITLE      cellulase (EC 3.2.1.4) precursor - Clostridium cellulovorans
ALTERNATE_NAMES
ORGANISM    endo-1,4-beta-glucanase
DATE        07-Jul-1997 #sequence_revision 12-Sep-1997 #text_change
ACCESSIONS JC5487; PC4333
REFERENCE   JC5487
#authors    Shewetta, S.A.; Ichi-ishi, A.; Park, J.S.; Liu, C.; Malburg,
               L.M.; Doi, R.H.
#journal    Gene (1996) 182:163-167
#title      Characterization of enfF, a gene for a non-cellulosomal
               Clostridium cellulovorans endoglucanase.
#cross-references MUID:97136706
#accession   JC5487
#molecule_type DNA
##residues    1-557 #label SHE1
##cross-references GB:U37056; NID:g1778708; PID:g1778709
##experimental_source strain ATCC 35296
#accession   PC4333
#molecule_type protein
##residues    30-37 #label SHE2
COMMENT      This enzyme plays a secondary role in cellulose degradation.
GENETICS
#gene        enfF
KEYWORDS     glycosidase; hydrolase; polysaccharide degradation
FEATURES
1-29         #domain signal sequence #status predicted #label SIG

```

```

30-557
SUMMARY      #length 557 #molecular-weight 60131 #checksum 6863

Query Match    24.9%; Score 850; DB 2; Length 557;
Best Local Similarity 41.3%; Pred. No. 4.96e-121;
Matches 133; Conservative 76; Mismatches 91; Indels 22; Gaps 13;

Db 63 LCDKRGPIQLRGASTHGLWFPFVVNNFAALSDNNSNVIRLAMYAEGGYATNPVS 122
QY 45 LVNERGEQVQLKGMSSHLQVGFVNTESMKWLDRDDGKITVFRAMYTSSGQYDDPS 104

Db 123 KOTVINGLYALANDMYVIVDWHMNPDPNASVSG--AQSPFENDISTLYPNKNIYEL 181
QY 105 KEKVEITVEAIDLGIYVIDWHILSDNDPN--IKKEAKDFFDESELYGDPNVIYEI 162

Db 182 CNEPENGSGVTNDATGWAQVKSATPIVQLLRKNENLIIVGNPFVSRQPDLAADNPI 241
QY 163 ANEPNGSD--VTWDN---QIKPYAEVPIVRNDPNNVIVGTGTSQDVHHAADNQL 216

Db 242 NDSNTMYSVHYFSGTNPISTVTDNRDNAMSNVRYALNHGAAVFATWGTSLATGTGPLY 301
QY 217 ADPNVYAFHYFAGTH--GQ-NL-RDQ-V-D--YALDQGAALFVSEWGTSAATGGGVFL 268

Db 302 AKADAWLDFLNGNNTSWCFNSISNKKERAAALNSITSLDPSDKLWADNELTSGQYVRA 361
QY 269 DEAQWIDFMDERNLSWANWSLTHKDESSAAL--MPGANPTGG--WTEAELSPSTFYRE 324

Db 362 RIKGAYATPVDPVT--NOPTAP 382
QY 325 KIRASATPPSDPTPPSPGEP 346

RESULT 14
ENTRY      C42360      #type complete
TITLE      cellulase (EC 3.2.1.4) E5 precursor - Thermomonospora fusca
ALTERNATE_NAMES
ORGANISM    endo-1,4-beta-glucanase
DATE        24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change
ACCESSIONS C42360
REFERENCE   N42360
#authors    Lao, G.; Ghangas, G.S.; Jung, E.D.; Wilson, D.B.
#journal    J. Bacteriol. (1991) 173:3397-3407
#title      DNA sequences of three beta-1,4-endoglucanase genes from
               Thermomonospora fusca.
#cross-references MUID:91258320
#accession   C42360
#status       preliminary
##residues    1-466 #label LAO
##cross-references GB:L01577; NID:g154693; PID:g154694
FUNCTION     #description hydrolysis of 1,4-beta-D-glucosidic linkages in
               beta-D-glucans such as cellulose and lichenin; can
               hydrolyze such linkages in beta-D-glucans that also contain
               1,3-linkages
               cellulose degradation
#pathway     cellulose degradation
CLASSIFICATION
KEYWORDS     #superfamily bacterial cellulose-binding domain homology
               glycosidase; hydrolase; polysaccharide degradation
FEATURES
38-137      #domain bacterial cellulose-binding domain homology
               #label BCB
SUMMARY      #length 466 #molecular-weight 49760 #checksum 9976

Query Match    23.5%; Score 805; DB 2; Length 466;
Best Local Similarity 39.9%; Pred. No. 4.16e-113;
Matches 122; Conservative 81; Mismatches 88; Indels 15; Gaps 11;

Db 165 VERYKQVCGTQLCDEHGNPVQLRGMSTHGQWFDHCLTSSDLALVDWKADIRLM 224
QY 32 VEEHQQLSISNGELVNERGEQVQLKGMSSHLQVGFVNTESMKWLDRDDGKITVFRAM 91

Db 225 YIQDGYETNPRGFTDRIDQLIDMATARGLYVIVDWHILTPGDPH--YNLDRAKTFPFAEA 283

```

Query Match	23.2%	Score 793;	DB 2;	Length 428;
Best Local Similarity	40.1%;	Pred. No. 5.33e-111;		
Matches	125;	Conservative	69;	Mismatches 106;
			Indels	12;
Gaps	10;			
Db	76	WGCEKFTYADTVASLAKDKNSITVRAAMGVOESGGVQLDPAGNKAKEVERVWDAATANDMY	135	
	I : :	: : :	: : :	: : :
QY	65	WYGO-FVNYESMKWLRRDNGCITVFRAAM-YTSSGGVIDDSPV-KEKVETVFAAIDLGIY	121	
	I : :	: : :	: : :	: : :
Db	136	AITGWHSIHAEN-N--RSEAIRFFQEMARKGNPNVIYEINEPL--OVWSNSTITKPYA	190	
	I I I I I	: : :	: : :	: : :
QY	122	VIIDWHILSDNDPNIYKEEAKDFDEMSELGYDPNVIIYEIANEPNGSDVTWDNQIKPYA	181	
	I I I I I	: : :	: : :	: : :
Db	191	EAVISAIRAITDPNLIIIVGPSQNVNDEASRDPINAKNIAITYTLHPYACTHGESLRNKAR	250	
	I : :	: : :	: : :	: : :
QY	182	EEVIPVRDRDPNNIIVGTGTSQDVHHRADNQLADPNMVAHFAYACTHGQNLDQVD	241	
	I : :	: : :	: : :	: : :
Db	251	QALNGCIAFVTEGWVINADGGGVNQOTETDAAWTFPMRDNNITLTQNWALNDKNEGASTY	310	
	I I I I I	: : :	: : :	: : :
QY	242	YALUDGAAILFVSWMGTSAATGDGCVFLDEQAVIDFMDERNLSWA-NWELTIHKDESSAAL	300	
	I I I I I	: : :	: : :	: : :

W P S R L H
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm

on: Fri Jun 4 10:01:01 1999; MasPar time 17.24 Seconds
 941.037 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-945-574-2
 Description: (1-574) from US08945574.pep
 Perfect Score: 4235
 Sequence: 1 MKWMSVMVLAVLVYFVA.....GNQVTGIAAQTTSKNKK 574

Scoring table: PAM 150
 Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot37
 1:swissprot

Statistics: Mean 52.851; Variance 97.011; scale 0.545

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	2570	60.7	1 GUNB_PAE1A	ENDOGLUCANASE B PRECUR	0.00e+00
2	691	16.3	1 GUNB_CLOCL	ENDOGLUCANASE B PRECUR	6.89e-117
3	673	15.9	1 GUNB_CLOCL	ENDOGLUCANASE D PRECUR	6.01e-113
4	660	15.6	1 GUNB_CLOCL	ENDOGLUCANASE E PRECUR	4.17e-110
5	614	14.5	1 GUNB_RUMAL	ENDOGLUCANASE A (EC 3.4.31e-100	4.31e-100
6	603	14.2	1 GUNB_RUMAL	ENDOGLUCANASE B PRECUR	1.05e-97
7	597	14.1	1 GUNB_RUMAL	ENDOGLUCANASE I PRECUR	2.08e-96
8	525	12.4	1 GUNB_NEOPA	ENDOGLUCANASE B PRECUR	6.43e-81
9	510	12.0	1 GUNB_CLOLO	ENDOGLUCANASE A PRECUR	1.02e-77
10	489	11.5	1 GUNB_CLOLO	ENDOGLUCANASE 1 (EC 3.2.96e-73	2.96e-73
11	479	11.3	1 GUNB_CLOCE	ENDOGLUCANASE A PRECUR	3.89e-71
12	422	10.0	1 GUNB_CLOCE	ENDOGLUCANASE D PRECUR	3.57e-59
13	412	9.7	1 GUNB_CLOCL	ENDOGLUCANASE H PRECUR	4.28e-57
14	139	3.3	1 GUNB_CRYFL	ENDOGLUCANASE I PRECUR	1.76e-05
15	141	3.3	1 GUN2_TRIRE	ENDOGLUCANASE EG-II PR	8.72e-06
16	127	3.0	1 YIS7_YEAST	HYPOTHETICAL 87.0 KD P	1.04e-03
17	118	2.8	1 GUNB_RUMFL	CELLULOSE DEGRADATION: HYDROLASE; GLYCOSIDASE; SIGNAL.	1.91e-02
18	119	2.8	1 CBPA_CLOCL	CELLULOSE BINDING PROT	1.39e-02
19	112	2.6	1 PRIM_ENTFA	DNA PRIMASE (EC 2.7.7)	1.22e-01
20	104	2.5	1 YUAB_BACSU	HYPOTHETICAL 19.3 KD P	1.29e+00
21	104	2.5	1 WAPA_BACST	WALL-ASSOCIATED PROTEI	1.29e+00
22	105	2.5	1 MDH_HALMA	MALATE DEHYDROGENASE (9.69e-01
23	105	2.5	1 YBG0_ECOLI	HYPOTHETICAL 39.5 KD P	9.69e-01

24	106	2.5	1 GLNA_HELPY	GLUTAMINE SYNTHETASE (7.25e-01
25	104	2.5	1 ESTS_DROVI	ESTERASE S PRECURSOR (1.29e+00
26	106	2.5	1 XINA_NEOPA	BIFUNCTIONAL ENDO-1,4-	7.25e-01
27	106	2.5	1 EF2K_CAEEL	ELONGATION FACTOR 2 KI	7.25e-01
28	104	2.5	1 YMD5_CAEEL	HYPOTHETICAL 101.3 KD	1.29e+00
29	100	2.4	1 IPYR_THELI	THERMOSTABLE INORGANIC	3.97e+00
30	100	2.4	1 PYHD_NPVHC	POLYHEDRIN MAJOR OCCL	3.97e+00
31	102	2.4	1 IN35_HUMAN	INTERFERON-INDUCED 35	2.27e+00
32	103	2.4	1 YYXA_BACSU	HYPOTHETICAL PROTEASE	1.72e+00
33	101	2.4	1 HLYB_PROMI	HEMOLYSIN ACTIVATOR PR	3.01e+00
34	103	2.4	1 GUNZ_CLOS	ENDOGLUCANASE 2 PRECUR	1.72e+00
35	103	2.4	1 BINC_EMENI	KINESIN-LIKE PROTEIN B	1.72e+00
36	99	2.3	1 RL4A_YEAST	60S RIBOSOMAL PROTEIN	5.23e+00
37	99	2.3	1 RL4B_YEAST	60S RIBOSOMAL PROTEIN	5.23e+00
38	99	2.3	1 GUN_BURSO	ENDOGLUCANASE PRECURSO	5.23e+00
39	99	2.3	1 CRTI_NEUCR	PHYTONE DEHYDROGENASE	5.23e+00
40	98	2.3	1 VACB_MYCPN	VACB PROTEIN HOMOLOG.	6.86e+00
41	99	2.3	1 DHG_ECOLI	GLUCOSE DEHYDROGENASE	5.23e+00
42	98	2.3	1 GUX2_CLOS	EXOGLUCANASE II PRECUR	6.86e+00
43	98	2.3	1 YG23_YEAST	HYPOTHETICAL 106.7 KD	6.86e+00
44	96	2.3	1 YJD9_YEAST	HYPOTHETICAL 191.5 KD	1.17e+01
45	98	2.3	1 RRPA_CVH22	RNA-DIRECTED RNA POLYM	6.86e+00

ALIGNMENTS

RESULT 1	STANDARD;	PRT;	566 AA.
ID GUNB_PAE1A			
AC P23550;			
DT 01-NOV-1991 (REL. 20, CREATED)			
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)			
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)			
DE ENDOGLUCANASE B PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE B)			
DE (CELLULOSE B) (EG-B).			
GN CELB.			
OS PAENIBACILLUS LAUTUS (BACILLUS LAUTUS).			
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;			
OC PAENIBACILLUS.			
RN (1)			
RP SEQUENCE FROM N.A.			
RC STRAIN-PL236;			
RX MEDLINE; 91033016.			
RA JORGENSEN P.L., HANSEN C.K.;			
RT "Multiple endo-beta-1,4-glucanase-encoding genes from Bacillus lautus			
RT PL236 and characterization of the celB gene.";			
RL GENE 93:55-60(1990).			
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC			
CC LINKAGES IN CELLULOSE.			
CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL			
CC HYDROLASES).			

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

DR EMBL; M3762; G142871; -			
DR EMBL; A28172; G905296; -			
DR PIR; JH0218; JH0218			
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.			
DR PFAM; PF00150; cellulase; 1.			
DR HSSP; P17901; 1EDG.			
KW CELLULOSE DEGRADATION: HYDROLASE; GLYCOSIDASE; SIGNAL.			
FT SIGNAL 1 30			
FT CHAIN 31 566			
FT ACT_SITE 177 177			
FT ACT_SITE 299 299			
SQ SEQUENCE 566 AA; 62621 MW; B6196713 CRC32;			

Query Match 60.7%; Score 2570; DB 1; Length 566;


```

OS CLOSTRIDIUM THERMOCELLUM.
OC BACTERIA; PHILICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.
RX MEDLINE; 89137992.
RX HALL J., HAZLEWOOD G.P., BARKER P.J., GILBERT H.J.;
RT "Conserved reiterated domains in Clostridium thermoCELLum
RT endoglucanases are not essential for catalytic activity.";
RL GENBANK 69-29-38(1988).
CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC GLUCANS.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSEOMES ENYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M22759; G144770; .
CC PIR; JTO347; CZCLEM.
CC DR PROSITE; PS00448; CLOS_CELLULOSEOME_RPT; 2.
CC DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
CC DR PFAM; PF00150; cellulase; 1.
CC DR PFAM; PF00404; celCC; 2.
CC DR PFAM; PF00657; Lipase_GDSL; 1.
CC DR HSP; P17901; LEDG.
CC CLOSTRIDIOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; REPEAT; SIGNAL.
CC CELLULOSE
CC SIGNAL 1 34
CC CHAIN 35 814 ENDOGLUCANASE E.
CC ACT_SITE 193 193 PROTON DONOR (BY SIMILARITY).
CC ACT_SITE 316 316 NUCLEOPHILE (BY SIMILARITY).
CC DOMAIN 415 474 2 X 24 AA APPROXIMATE REPEATS.
CC REPEAT 415 438 1.
CC REPEAT 451 474 2.
CC SEQUENCE 814 AA; 90244 MW; AE61A167 CRC32;
CC -----
Query Match 15.6%; Score 660; DB 1; Length 814;
Best Local Similarity 37.1%; Pred. No. 4.17e-110;
Matches 104; Conservative 72; Mismatches 92; Indels 12; Gaps 10;

Db 61 VKEIKGWNGLTDA--PTETAGNPRRTTKAMIEKVRMGFNARVPVVTWDTHTIGAPD 118
QY 39 VRDQPGQNNLGNFTFDAYQGDQGTAGNPRVTRIELIADIEGYKSRIPVTWENRIGAPD 98
Db 119 YKIDEAWLNREEVNVVLDCGMYAIIHLNHDN-TWIIPTYANEQRSKEKLVKVVWQIAT 177
QY 99 YPIDQFELNRDVEVQWALSEEDLYVMNLNHDHSLWLYEMEHYNGVMARYSLWLSQLSN 158
Db 178 RFKDYDDLHLETFETMNEPRVSGPMWGGTYENRD-VINRFLNAVNTIRASGNNDKRF 236
QY 159 HFNDYPTPKLFESVNEPK-F-SQ-NW-GEIRENHALLDLNLTNTFFELVRQSGQNDIRP 214
Db 237 ILVPTNAATGLDVALNDL--VINP-NDRSRVIVSHAYSPFFAMVNGTSGWSDYDKAS 293
QY 215 LVLPMTATSSQPLNNLYQTDIKLDNPLNIATVHYGYFWPFSVIAGYRFEED-SKRE 273
Db 294 LTSELDAILYNRVKNRGAIVIEGEGTIDKNNLSRYAAHE 333
QY 274 IIFEDRVHHTFVARGIPVVLGFEGLLGFDKHTGVIOQGE 313
RESULT 5

```

ID CUNA_RUMAL STANDARD; PRT; 364 AA.
 AC P23660;
 DT 01-NOV-1991 (REL. 20, CREATED)
 DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE ENDOGLUCANASE A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULASE)
 DE (EGA).
 GN CELA.
 OS RUMINOCOCCUS ALBUS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
 OC RUMINOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-14.
 RC STRAIN-SY3;
 RX MEDLINE; 91066833.
 RA POOLE D.M., HAZLEWOOD G.P., LAURIE J.I., BARKER P.J., GILBERT H.J.;
 RT "Nucleotide sequence of the Ruminococcus albus SY3 endoglucanase
 RT genes cels and celsB".
 RL MOL. GEN. GENET. 223:217-223(1990).
 CC -1- FUNCTION: HYDROLYSES BOTH CARBOXYMETHYLCELLULOSE AND XYLAN.
 CC PROBABLY HAS A ROLE IN HYDROLYZING OLIGOSACCHARIDES DERIVED
 CC FROM CELLULOSE, WHICH ARE TRANSPORTED ACROSS THE CELL WALL.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC
 CC LINKAGES IN XYLANS.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X54931; G45964; -
 CC PIR; S12017; S12017.
 CC PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 CC PFAM; PF00150; cellulase; 1.
 CC HSSP; P17901; 1EDG.
 CC CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; XYLAN DEGRADATION.
 CC ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
 CC ACT_SITE 293 293 NUCLEOPHILE (BY SIMILARITY).
 CC SEQUENCE 364 AA; 41218 MW; 680A4052 CRC32;
 CC -----
 CC Query Match 14.5%; Score 614; DB 1; Length 364;
 CC Best Local Similarity 35.8%; Pred. No. 4.31e-100;
 CC Matches 101; Conservative 68; Mismatches 94; Indels 19; Gaps 16;
 Db 21 EVRDISAMELVGEMKTKGNLGNLSLDATGAPGNASEVWNGPKTKEMIDAVYNGKGFVIR 80
 QY 29 DVKTLDTQSVRDMQPGWNLGNFTDVG-Q-Q--ETAWGNPRVTRIELIERADEGYSIR 84
 Db 81 IPTVWGHVGDADPKIDDEWIARVQWVAYDDGAYVINSHHEEDW-RIPDNEHI-D 138
 QY 85 IPTWENRIGAPPDYPDPQFLNRVDEVQWALEEDLYVMNLHHD-SWLWIYEMEHYN 143
 Db 139 AVEKTAAIKWVAERPKDYGDHLIFGLNEPRVKVGSQEWNGGTGEGRCV-DRLNKTFF 197
 QY 144 GVMKYRSLNEQLSNHFKDPTKLMFESVNEPKF--S-QNW-GEIRENHALLDLDLTVF 199
 Db 198 LDTVRATGGNEXR-LLL--MTIYASSM-SNVIKDTAIPEDDHIGFSIHAYPYATYN 253
 QY 200 FEIVRQSGGQNDIRPLVPLMTETATSOPLNNLYQTIDKLDPLNIATVHYGFWPFVSN 259
 Db 254 ANADWELFWHDGSDGELVSLMNLKENYLDKQIPVIITEYG 295
 QY 260 I-AGYTRFE-EDSKR-EIETTFORVHHTFVARGIPVVLGEFG 298

AC Q12647;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE B PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE B)
DE (CELLULOSE B).
GN CELB.
OS NEOCALLIMASTIX PATRICIARUM (RUMEN FUNGUS).
OS EUKARYOTA; FUNGI; CHYTRIDIOMYCOTA; CHYTRIDIOMYCETES;
OC NEOCALLIMASTICALES; NEOCALLIMASTICACEAE; NEOCALLIMASTIX.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94128068.
RT ZHOU L., XUE G., ORPIN C.G., BLACK G.W., GILBERT H.J., HAZLEWOOD G.P.;
RA "Intronless cells from the anaerobic fungus Neocallimastix patriciarum
RT encodes a modular family A endoglucanase.";
RL BIOCHEM. J. 297:359-364(1994).
CC -1- FUNCTION: RATE OF HYDROLYSIS OF CELLULO-OLIGOSACCHARIDES INCREASED
CC WITH INCREASING CHAIN LENGTH FROM CELLULOSE TO CELLOPENTAOSE.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z31364; G467687;
DR PROSITE; PS00659; GLYCOSYL-HYDROL_F5; 1.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PFAM; PF00150; cellulase; 1.
DR HSSP; P17901; LEDG.
KW HYDROLASE; GLYCOSIDASE; REPEAT; SIGNAL.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 473 ENDOGLUCANASE B.
FT DOMAIN 18 367 CATALYTIC.
FT DOMAIN 365 387 LINKER.
FT DOMAIN 376 381 POLY-THR.
FT DOMAIN 390 473 2 X 39 AA APPROXIMATE REPEATS.
FT REPEAT 390 427 1.
FT REPEAT 435 473 2.
FT ACT_SITE 173 173 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 295 295 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 473 AA; 53070 MW; 929CE396 CRC32;

Query Match 12.48; Score 525; DB 1; Length 473;
Best Local Similarity 29.68; Pred. No. 6.43e-81;
Matches 104; Conservative 103; Mismatches 118; Indels 26; Gaps 20;

Db 1 MKFLNTFSLSLAIGSKAMKNISS-KELVKDLTIGSLGNTLDATCFETLDYKNKQIA- 58
QY 1 MKWKSVMVLAVLVVSVAPAVSSANEDYKTLDIQ-SYVRDQPG-WN-LGNTFDVAGQ 57

Db 59 SETCWGNKVTQELLYKSLDGLGFNTPRIPPTWTSGHFNADPYKINDQMKRVHIVYAI 118
QY 58 DETAWGNPRVTRIELIERIADGKYKSIPTWENRIGGADYPIDPQFLNRVDEVQWAL 117

Db 119 NTGGYATLNTHTWTHAFQK-NLESAKILVAIKQIAAEADYDEHLIFECMNEPK 176
QY 118 EEDLYVMINLHDSWLIWYEMEHNYNGVMAYKSLWQLSNHFKDYPTKLMFESVNEP-K 176

Db 177 VGDPAAEWNGDYGWN-FVNEEMNDLFVKTRATGTGNNALRHLMIPTVAAACINDGAINN-F 234
QY 177 FSQ--NW-GEIRENHALLDLDNTVFEIVRQSGQNDIRPLVLPMTETATSPQLLNLY 233

Db 235 K-FPSGDD-KVIVSLHSYSPYNFALNNGAGASINFYDGS--EIDWAMNTINSKFSIGIP 290
QY 234 QTDKLDPPNLIAIVHYHGFVSPVNT-AG-YTFEEDSKREIETEDRVHHTFVARGIP 291

RESULT 8
ID GUNB_NEOFA STANDARD; PRT; 473 AA.

AC Q12647;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE B PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE B)
DE (CELLULOSE B).
GN CELB.
OS NEOCALLIMASTIX PATRICIARUM (RUMEN FUNGUS).
OS EUKARYOTA; FUNGI; CHYTRIDIOMYCOTA; CHYTRIDIOMYCETES;
OC NEOCALLIMASTICALES; NEOCALLIMASTICACEAE; NEOCALLIMASTIX.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94128068.
RT ZHOU L., XUE G., ORPIN C.G., BLACK G.W., GILBERT H.J., HAZLEWOOD G.P.;
RA "Intronless cells from the anaerobic fungus Neocallimastix patriciarum
RT encodes a modular family A endoglucanase.";
RL BIOCHEM. J. 297:359-364(1994).
CC -1- FUNCTION: RATE OF HYDROLYSIS OF CELLULO-OLIGOSACCHARIDES INCREASED
CC WITH INCREASING CHAIN LENGTH FROM CELLULOSE TO CELLOPENTAOSE.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z31364; G467687;
DR PROSITE; PS00659; GLYCOSYL-HYDROL_F5; 1.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PFAM; PF00150; cellulase; 1.
DR HSSP; P17901; LEDG.
KW HYDROLASE; GLYCOSIDASE; REPEAT; SIGNAL.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 473 ENDOGLUCANASE B.
FT DOMAIN 18 367 CATALYTIC.
FT DOMAIN 365 387 LINKER.
FT DOMAIN 376 381 POLY-THR.
FT DOMAIN 390 473 2 X 39 AA APPROXIMATE REPEATS.
FT REPEAT 390 427 1.
FT REPEAT 435 473 2.
FT ACT_SITE 173 173 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 295 295 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 473 AA; 53070 MW; 929CE396 CRC32;

Query Match 12.48; Score 525; DB 1; Length 473;
Best Local Similarity 29.68; Pred. No. 6.43e-81;
Matches 104; Conservative 103; Mismatches 118; Indels 26; Gaps 20;

Db 1 MKFLNTFSLSLAIGSKAMKNISS-KELVKDLTIGSLGNTLDATCFETLDYKNKQIA- 58
QY 1 MKWKSVMVLAVLVVSVAPAVSSANEDYKTLDIQ-SYVRDQPG-WN-LGNTFDVAGQ 57

Db 59 SETCWGNKVTQELLYKSLDGLGFNTPRIPPTWTSGHFNADPYKINDQMKRVHIVYAI 118
QY 58 DETAWGNPRVTRIELIERIADGKYKSIPTWENRIGGADYPIDPQFLNRVDEVQWAL 117

Db 119 NTGGYATLNTHTWTHAFQK-NLESAKILVAIKQIAAEADYDEHLIFECMNEPK 176
QY 118 EEDLYVMINLHDSWLIWYEMEHNYNGVMAYKSLWQLSNHFKDYPTKLMFESVNEP-K 176

Db 177 VGDPAAEWNGDYGWN-FVNEEMNDLFVKTRATGTGNNALRHLMIPTVAAACINDGAINN-F 234
QY 177 FSQ--NW-GEIRENHALLDLDNTVFEIVRQSGQNDIRPLVLPMTETATSPQLLNLY 233

Db 235 K-FPSGDD-KVIVSLHSYSPYNFALNNGAGASINFYDGS--EIDWAMNTINSKFSIGIP 290
QY 234 QTDKLDPPNLIAIVHYHGFVSPVNT-AG-YTFEEDSKREIETEDRVHHTFVARGIP 291

RESULT 8
ID GUNB_NEOFA STANDARD; PRT; 473 AA.

AC Q12647;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE B PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE B)
DE (CELLULOSE B).
GN CELB.
OS NEOCALLIMASTIX PATRICIARUM (RUMEN FUNGUS).
OS EUKARYOTA; FUNGI; CHYTRIDIOMYCOTA; CHYTRIDIOMYCETES;
OC NEOCALLIMASTICALES; NEOCALLIMASTICACEAE; NEOCALLIMASTIX.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94128068.
RT ZHOU L., XUE G., ORPIN C.G., BLACK G.W., GILBERT H.J., HAZLEWOOD G.P.;
RA "Intronless cells from the anaerobic fungus Neocallimastix patriciarum
RT encodes a modular family A endoglucanase.";
RL BIOCHEM. J. 297:359-364(1994).
CC -1- FUNCTION: RATE OF HYDROLYSIS OF CELLULO-OLIGOSACCHARIDES INCREASED
CC WITH INCREASING CHAIN LENGTH FROM CELLULOSE TO CELLOPENTAOSE.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z31364; G467687;
DR PROSITE; PS00659; GLYCOSYL-HYDROL_F5; 1.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PFAM; PF00150; cellulase; 1.
DR HSSP; P17901; LEDG.
KW HYDROLASE; GLYCOSIDASE; REPEAT; SIGNAL.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 473 ENDOGLUCANASE B.
FT DOMAIN 18 367 CATALYTIC.
FT DOMAIN 365 387 LINKER.
FT DOMAIN 376 381 POLY-THR.
FT DOMAIN 390 473 2 X 39 AA APPROXIMATE REPEATS.
FT REPEAT 390 427 1.
FT REPEAT 435 473 2.
FT ACT_SITE 173 173 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 295 295 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 473 AA; 5

Db 291 VIIGEGAM--NRNN--ED-DRERAEYIKKATSIGVPCVINDG-YFE 334
 QY 292 VVIGEGFLGDKHTGVIOQGEKLFEXYLIHHLNERDITHMLWDNGQHFN 342

RESULT 9
 ID GUNA_CLOLO STANDARD; PRT; 517 AA.
 AC P54937;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE ENDOGLUCANASE A PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE A)
 DE (CELLULOSE A).
 GN CELA.

OS CLOSTRIDIUM LONGISPORUM.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
 OC CLOSTRIDIUM.

RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN-ATCC 49440;
 RX MEDLINE; 94172316.

RA MITTENDORF V., THOMSON J.A.;
 RT "Cloning of an endo-(1->4)-beta-glucanase gene, celsa, from the rumen
 bacterium Clostridium sp. ('C. longisporum') and characterization of
 its product, Celsa, in Escherichia coli.";
 RL J. GEN. MICROBIOL. 139:3233-3242(1993).

CC -!- FUNCTION: THE PH AND TEMPERATURE OPTIMA OF CELA ARE 4.8 AND 43
 DEGREES CELSIUS, RESPECTIVELY. IT HYDROLYSES BARLEY BETA-GLUCAN,
 LICHENAN, CARBOXYMETHYLCELLULOSE AND XYLAN. IT SHOWS PREFERENTIAL
 ACTIVITY AGAINST THE LARGER CELLOOLIGOSACCHARIDES (CELLOHEXAOSE
 AND CELLOPENTAOSE); CELLOTRAOSE IS THE SMALLEST SUBSTRATE
 DEGRADED COMPLETELY.

CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 LINKAGES IN CELLULOSE.

CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 HYDROLASES).

CC -!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
 (CBD).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; L02868; G144755; -
 DR PROSITE; P500659; GLYCOSYL_HYDROL_F5; 1.
 DR PFAM; PF00150; cellulase; 1.
 DR PFAM; PF00553; CBD_1; 1.
 DR HSP; P17901; 1EDG.
 DR CELLULOSE DEGRADATION: HYDROLASE; GLYCOSIDASE; SIGNAL.
 FT SIGNAL 1 25
 FT CHAIN 26 517
 FT DOMAIN 26 ?
 FT ACT_SITE 421 517
 FT ACT_SITE 185 185
 FT ACT_SITE 309 309
 FT ACT_SITE 517 AA; 57660 MW; 8DEC293D CRC32;

Query Match 12.0%; Score 510; DB 1; Length 517;
 Best Local Similarity 31.8%; Pred. No. 1.02e-77;
 Matches 91; Conservative 85; Mismatches 89; Indels 21; Gaps 17;

Db 73 EFGWGNPVTKAMIDKIKNAGFKTIRIPITWGEHLGN-N-KLNEWVRKVEVDYCIA 130

QY 59 ETAWGNPRVTRIELIERIADEGYKSIRIPVTWENRIGGAPDYPIDPQFLNRVDEVQWALE 118

Db 131 DDLVILNTHHEG-NWVITYAKESVTPKLTLTWQISEAFKDYDDHLIFETLNPRLE 189

QY 119 EDLYVINHLHSDWLIYEMEHNYNGVMARYSLWQLSNHFKDPTKLMFESVNEPKF- 177

Db 190 GTPYEWTTGGTSESD-VVKNYNAALLESIRKTGGNNLSRAVMMPTYAASGSSTTMD-FK 247
 QY 178 -SQ-NW-GEIRENHALLDLNTVFEIVRSGSQNDIRPLVLTMTATSQPLNLYQ 234
 Db 248 -VP--DDKNVIASVHAYSPYFFAMDTSSNSVNTWGSSYDKYSLDVELDYLNTFKSGVP 304
 QY 235 TIDKLDPNLIAVHYGFWPFSVNIAGYT-R-FEESKRELIET-FDRVHTFTVARGIP 291
 Db 305 VVIGEGF--SINKNNTS-SRAE-LA--EYVTAQKRGIPCVWMDN 344
 QY 292 VVLGEGLLGDKHTGVIOQGEKLFEXYLIHHLNERDITHMLWDN 337

RESULT 10
 ID GUNL_BUTFI STANDARD; PRT; 547 AA.
 AC P20847;
 DT 01-FEB-1991 (REL. 17, CREATED)
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE ENDOGLUCANASE 1 (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE).
 GN END1.

OS BUTYRIVIBRIO FIBRISOLVENS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
 OC BUTYRIVIBRIO.

RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-H17C;
 RX MEDLINE; 90136507.

RA BERGER E., JONES W.A., JONES D.T., WOODS D.R.;
 RT "Cloning and sequencing of an endoglucanase (endl) gene from
 Butyrivibrio fibrisolvens H17c.";

RL MOL. GEN. GENET. 219:193-198(1989).

CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 LINKAGES IN CELLULOSE.

CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 HYDROLASES).

CC -!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
 (CBD).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; X17538; G39473; -
 DR PIR; JQ0356; JQ0356.
 DR PROSITE; P500561; CBD_BACTERIAL; 1.
 DR PROSITE; P500659; GLYCOSYL_HYDROL_F5; 1.
 DR PFAM; PF00150; cellulase; 1.
 DR PFAM; PF00553; CBD_1; 1.
 DR HSP; P17901; 1EDG.
 DR CELLULOSE DEGRADATION: HYDROLASE; GLYCOSIDASE.
 FT ACT_SITE 189 189
 FT ACT_SITE 321 321
 FT ACT_SITE 451 547
 FT DOMAIN 547 AA; 61078 MW; F318ABC3 CRC32;

Query Match 11.5%; Score 489; DB 1; Length 547;
 Best Local Similarity 32.9%; Pred. No. 2.96e-73;
 Matches 103; Conservative 84; Mismatches 94; Indels 32; Gaps 19;

Db 15 LALFVFMIPAIPKVSAAAGTDRS-ATQV-VSDMRVGNWGNISLDSFGSQSYNFPYTSLE 72

QY 10 LAV-VLVWSPVAPVSSANEDVKTLDIQSVVRDQPCWNLGNFTDVGQ-----DE 59

Db 73 TYGNCNPATKALIDEVAKAGFNIRIPVSGQYTTGS-DYQI-PDFVMNRKVEVDYCI 130

QY 60 TAWGNPRVTRIELIERIADEGYKSIRIPVTWENRIGGAPDYPIDPQF-LNRVDEVQWALE 118

WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
-1- THE C-TERMINUS (AA 411-475) MAY PLAY A ROLE IN ORGANIZING THE
CELLULOSE COMPLEX.
-1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
HYDROLASES).

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

[illegible]

[illegible]

```

Db 137 IWEOIAQRFKSGKENIVFEILNEP---H--GNITDSQ---INDMNKRILNIIRKT---NP 183
      :|||::|||::|:|||::|||::|:|||::|:|||::|:|||::|:|||::|:|||::|
QY 152 LWEQLSNHFEDYPTKLMEFSVNEPKFQNWGEIRENHALLDLNTVFVEIQRSGGND 211
      :|||::|||::|:|||::|:|||::|:|||::|:|||::|:|||::|:|||::|
Db 186 TRN-VI--ICAGWYSY-NSLSQ-LeIPNDPNLIATFHYYDPYSTHOWOG-TWGTKNDM 239
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 212 IRPLVLPTMETATSQPLLNLNYQTIDKDDPNLIATVHYGYFWPFVSVIAGTYRFEEDSK 271
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 240 DATAMFNHV-KWSDKNNIPVYLGEVGMVSHSDRTSAV 277
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 272 RELIEFDVRVHTFVARG-IPVVLGEGLLGFKDHTGI 309
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 13
ID GUNH_CLOTW STANDARD; PRT; 900 AA.
AC P6218;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE ENDOGLUCANASE H PRECURSOR (EC 3.2.1.4) (EGH) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE H).
DE CE(LL)ULOSE H).
GN GN
OS CLOSTRIDIUM THERMOCELLUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
[1]
RN RP
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB 10682;
RX MEDLINE; 90323606.
RA YAGUEE E., BEGUIN P., AUBERT J.-P.;
RT "Nucleotide sequence and deletion analysis of the cellulase-encoding gene celH of Clostridium thermocellum.";
RL GENE 89:61-67(1990).
CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-GLUCANS.
CC CC
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.
CC CC
CC -1- DOMAIN: A 24 RESIDUOUS DOMAIN IS REPEATED TWICE IN THIS ENZYME AS WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC CC
CC -1- SIMILARITY: THE N-TERMINAL PART BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC CC
CC -1- SIMILARITY: THE C-TERMINAL PART BELONGS TO FAMILY 26 OF GLYCOSYL HYDROLASES.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcements or send an email to license@isb-sib.ch).
CC CC
EMBL; M31903; G144774; -.
DR PIR; JH0157; JH0157.
DR DR
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00448; CLOS_CELLULOSE_RFT; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF00404; celCC; 2.
DR HSP; P07985; ICEC.
KW CELLULOSE DEGRADATION: HYDROLASE; GLYCOSIDASE; REPEAT; SIGNAL.
FT SIGNAL 1 44
FT CHAIN 45 900 ENDOGLUCANASE H.
FT DOMAIN 45 630 CATALYTIC (BY SIMILARITY).
FT DOMAIN 631 654 PROTHR-RICH (LINKER).
FT DOMAIN 655 900 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 460 460 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 565 565 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 833 895 2 X 24 AA APPROXIMATE REPEATS.
FT REPEAT 833 856 1.
FT REPEAT 872 895 2.

```

```

Query Match      3.3%; Score 139; DB 1; Length 341;
Best Local Similarity 20.6%; Pred. No. 1.76e-05;
Matches 22; Conservative 31; Mismatches 52; Indels 2; Gaps 2;

Db      62 EQVGHFTADGANLFRLPAGHQYLVGNNGQASTSLAPDFAQYDALYQAVISGAYAIIDVH 121
      | : : : | : | : | : : : | : : | : | : : | : | : | : | : | : | : |
QY      70 ELERIADEGYSKIRIPVTWENRIGGA-PDYPIDPQFLNRVDVOWALEEDLYVMINLH 128
      | : : : | : | : | : : : | : : | : | : : | : | : | : | : | : | : |

Db      122 NYAR-WNGAIIQGGGSPQDFANLTLLATKVTNSDPNVIFGLMNEP 167
      | : : : | : | : | : : | : | : | : | : | : | : | : | : | : | : | : |
QY      129 HDSWLWYIEHHNYNGVMAYKIRSLWEQLSNHFKDYPFKLMFESVNEP 175
      | : : : | : | : | : : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 15
ID  GUN2_TRINE      STANDARD;      PRT;      418 AA.
AC  P07982;
DT  01-AUG-1988 (REL. 08, CREATED)
DT  01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
DT  15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE  ENDOGLUCANASE EG-II PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE  (CELLULOSE).
GS  EGL2 OR EGLI1.
GN  TRICHODERMA RESEI (HYPOCREA JECORINA).
OC  EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC  HYPOCREALES; HYPOCREACEAE; HYPOCREA.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-VTT-D-80133;
RX  MEDLINE: 88255850.
RA  SALOHEIMO M., LEHTOVAARA P., PENTTILA M., TEERI T.T., STAHLBERG J.,
RA  JOHANSSON G., PETERSSON G., CLAYSENS M., TOMME P., KNOWLES J.K.C.;
RT  "EgIII, a new endoglucanase from Trichoderma reesei: the
RT  characterization of both gene and enzyme.";
RL  GENE 63:11-21(1988).
RP  [2]
RP  ACTIVE SITE GLU-350.
RX  MEDLINE: 93131031.
RA  MACARRON R., VAN BEEUMEN J., HENRISSAT B., DE LA MATA I.,
RA  CLAEYSENS M.;
RT  "Identification of an essential glutamate residue in the active site
RL  of endoglucanase III from Trichoderma reesei.";
RL  FEBS LETT. 316:137-140(1993).
CC  -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC  GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC  (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC  (2) EXOCELLULOBIODHROLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE;
CC  FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC  (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC  SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC  -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC  LINKAGES IN CELLULOSE.
CC  -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC  HYDROLASES).
CC  -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC  -1- CAUTION: WAS ORIGINALLY CALLED ENDOGLUCANASE EG-III.
CC  -----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  at the European Bioinformatics Institute. There are no restrictions on
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announcement
CC  or send an email to license@isb-sib.ch).

```

```

FT SIGNAL 1 21
FT CHAIN 22 418
FT DOMAIN 22 57
FT DOMAIN 58 91
FT DOMAIN 92 418
FT MOD_RES 22 22
FT CARBOHYD 124 124
FT DISULFID 29 46
FT DISULFID 40 56
FT ACT_SITE 239 239
FT ACT_SITE 350 350
SQ SEQUENCE 418 AA; 44227 MW; 7C18782A CRC32;

Query Match 3.3%; Score 141; DB 1; Length 418;
Best Local Similarity 22.8%; Pred. No. 8.72e-06;
Matches 23; Conservative 30; Mismatches 46; Indels 2; Gaps 2;

Db 142 VNEDGMTIFRLPVGWQYLVNNGNLDSTISKYDQLVQGLSLGAYCIVDIHNYAR-W 200
QY 75 IADGYSIRIPVTWENRIGGAPDYPDPQLNRYDEVVQWALEEDLYVMINLHDSWLW 134

Db 201 NGGIGGGPTNAOFTSLWSQLASKYAS-QSRVWFGINNEP 240
QY 135 IYEMEHNYNGVMKYRSLWEQLSNHFKDYP TKLMEESVNEP 175

```

Search completed: Fri Jun 4 10:01:32 1999
Job time : 31 secs.

(TM)

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	3419	100.0	467	19 W00382	Bacillus cellulase BC	1.17e-253
2	3419	100.0	467	20 W05731	Cellulase.	1.17e-253
3	3249	95.0	462	32 W37433	Classed alkaline endog.	2.68e-240
4	2832	82.8	409	1 P18433	Sequence of alkaline	1.47e-207
5	2832	82.3	411	21 W2379	P300-CelB fusion cons	3.79e-206
6	2781	81.3	400	25 W3601	Bacillus agaradherens	1.47e-203
7	2781	81.3	400	32 W5741	Bacillus agaradherens	1.47e-203
8	2781	81.3	400	25 W2521	Bacillus agaradherens	1.47e-203
9	2772	81.1	410	21 W2378	P300-CelB fusion cons	7.46e-203
10	2701	79.0	411	21 W12381	P300-CelB fusion cons	2.76e-197
11	2691	78.7	412	21 W12380	P300-CelB fusion cons	1.68e-196
12	1649	48.2	499	8 R42122	NK-1 cellulase.	4.80e-115
13	1561	45.7	551	24 W8790	Corrected Bacillus la	3.38e-108
14	1339	39.2	531	20 W0150	60 kD endoglucanase,	5.84e-91
15	1329	38.9	532	3 R13229	Endoglucanase encoded	3.48e-90
16	862	25.2	1010	28 W34989	Teredinibacter endogl	3.48e-54

[illegible]

CONFIDENTIAL

QY 1 MKKITTIFAVLLMTLALFSIGNTTAADDYSVVEEHGQLSISNGELVNERGEQVQLKGMSS 60
 Db 61 hglwygqfnyesmkwlrddwgitvfraamytssggiddpsvkvetveaaidlgi 120
 QY 61 HGLQWYGQFVNYESMKWLRRDDWGITVFRAMYTSSGGYIDDPVSKVKETVEAAIDLGI 120
 Db 121 yviidwhillsndpnnykeakdfidenseiygdyvpnyielanepngsdvtdnqikpy 180
 QY 121 YVIIDWHILLSNDPNNYKEAKDFIDENSELYGDPNVIYEIANEPNGSDVTWQNIKPY 180
 Db 181 aeevipvirdndpnnyivgtgtsqdvhaadnqladpnmvmyafhfyagthgqnlrdqv 240
 QY 181 AEEVIPVIRDNDPNNVIVGTGTSQDVVHAADNQLADPNVMYAFHFYAGTHGQNLRDQV 240
 Db 241 dyaldqgaafvsewgttsaagdgvgvfdieaqvwidfndernlswanwslthkdessaal 300
 QY 241 DYALDQGAAFVSEWGTSAAGDGGVFDIEAQVWIDFNDERNLSWANWSLTHKDESSAAL 300
 Db 301 mpqantpggtwteaelspgstfvrekiresasippsdtpspdpgepdpptpsdpg 360
 QY 301 MPGANPTGGTWEAEELSPGSTFVREKIREASIPPSDTPSPDPGEPDPTPPSPDG 360
 Db 361 eypawdsnqiytnelviyhgqlwqakwtqngqepdpgpweplksdpdsgedpdpssd 420
 QY 361 EYPAWDSNQIYTNELVIYHNGQLWQAKWWTQNGEPDPGPWEPLKSDPDSDGEPDPTPPSD 420
 Db 421 pgeypawdsnqiytnelviyhgqlwqakwtqngqepdpgpwepln 467
 QY 421 PGEYPAWDSNQIYTNELVIYHNGQLWQAKWWTQNGEPDPGPWEPLN 467

RESULT 2
 ID W05731 standard; Protein; 467 AA.
 AC W05731;
 DT 29-MAR-1997 (first entry)
 DE Cellulase.
 KW Cellulase; Bacillus; alkalophilic bacterium; surfactant;
 KS stonewashing; biopolishing; fabric softener; depilling.
 OS Bacillus sp. strain CBS 670.93.
 FH Key
 FT Key
 FT Location/Qualifiers
 FT 1..26
 FT /note= "Signal peptide"
 FT EP-739982-A1
 PD 30-OCT-1996
 PF 30-OCT-1996
 PR 30-APR-1995; EP-201115
 PA (GEMV) GENENCOR INC.
 PI Lenting HBM, Maurer KH, Van Beckhoven RFWC;
 PI Van Solingen P, Weiss A;
 DR WPI; 96-478746/48.
 DR N-PSDB; T40008.
 PT Cellulase isolated from novel Bacillus species - useful in detergent compns, fabric softeners and de-pilling compns, exhibits reduced adsorption of enzyme to fabric
 CC Claim 6; Page 12-14; 17pp; English.
 CC The sequence represents a cellulase from alkalophilic Bacillus sp. CBS 670.93, and has been isolated in recombinant form by screening of Escherichia coli clones in plasmid pTZ18R. The enzyme shows a delta-REM of at least 4 units, preferably at least 5 units, in the Anti Redispersion Test, strong depilling activity, fibre damage of less than 0.05 mU in the Fibre Damage Test, and adsorption of less than 15% in the Adsorption Test. The cellulase may be used in surfactant compositions, and in stonewashing, biopolishing, fabric softener and depilling treatment compositions (claimed). The enzyme does not accumulate on the fabric after repeated laundry washing, and therefore has no effect on fabric tensile strength.
 SQ Sequence 467 AA;

Query Match 100.0%; Score 3419; DB 20; Length 467;
 Best Local Similarity 100.0%; Pred. No. 1.17e-253;
 Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mkkittifavllmtlalfsignttaaddysvveehgqlsngelvnrggeqvlkgmss 60

QY 1 MKKITTIFAVLLMTLALFSIGNTTAADDYSVVEEHGQLSISNGELVNERGEQVQLKGMSS 60
 Db 61 hglwygqfnyesmkwlrddwgitvfraamytssggiddpsvkvetveaaidlgi 120
 QY 61 HGLQWYGQFVNYESMKWLRRDDWGITVFRAMYTSSGGYIDDPVSKVKETVEAAIDLGI 120
 Db 121 yviidwhillsndpnnykeakdfidenseiygdyvpnyielanepngsdvtdnqikpy 180
 QY 121 YVIIDWHILLSNDPNNYKEAKDFIDENSELYGDPNVIYEIANEPNGSDVTWQNIKPY 180
 Db 181 aeevipvirdndpnnyivgtgtsqdvhaadnqladpnmvmyafhfyagthgqnlrdqv 240
 QY 181 AEEVIPVIRDNDPNNVIVGTGTSQDVVHAADNQLADPNVMYAFHFYAGTHGQNLRDQV 240
 Db 241 dyaldqgaafvsewgttsaagdgvgvfdieaqvwidfndernlswanwslthkdessaal 300
 QY 241 DYALDQGAAFVSEWGTSAAGDGGVFDIEAQVWIDFNDERNLSWANWSLTHKDESSAAL 300
 Db 301 mpqantpggtwteaelspgstfvrekiresasippsdtpspdpgepdpptpsdpg 360
 QY 301 MPGANPTGGTWEAEELSPGSTFVREKIREASIPPSDTPSPDPGEPDPTPPSPDG 360
 Db 361 eypawdsnqiytnelviyhgqlwqakwtqngqepdpgpweplksdpdsgedpdpssd 420
 QY 361 EYPAWDSNQIYTNELVIYHNGQLWQAKWWTQNGEPDPGPWEPLKSDPDSDGEPDPTPPSD 420
 Db 421 pgeypawdsnqiytnelviyhgqlwqakwtqngqepdpgpwepln 467
 QY 421 PGEYPAWDSNQIYTNELVIYHNGQLWQAKWWTQNGEPDPGPWEPLN 467

RESULT 3
 ID W57433 standard; Protein; 462 AA.
 AC W57433;
 DT 01-SEP-1998 (first entry)
 DE Cloned alkaline endoglucanase protein sequence.
 KW Starch; liquefaction; sweetener; enzyme hybrid; endoglucanase; enzyme;
 KS cellulose binding domain; CBD; starch processing; alpha-amylase;
 KW saccharification.
 OS Bacillus agaradherens.
 FH Key
 FT Key
 FT Location/Qualifiers
 FT Peptide
 FT 1..26
 FT /note= "signal peptide"
 FT Protein
 FT 27..326
 FT /note= "endoglucanase enzyme"
 FT Peptide
 FT 327..354
 FT /note= "to be used as a linker sequence"
 FT Domain
 FT 355..400
 FT /note= "cellulose binding domain"
 FT Peptide
 FT 401..416
 FT /note= "to be used as a linker sequence"
 FT Domain
 FT 417..462
 FT /note= "cellulose binding domain"

FN W057433-A1.
 PD 30-OCT-1998
 PF 30-OCT-1998
 PR 11-OCT-1996; DK-001130
 PA (NOVO) NOVO-NORDISK-AS.
 PI Bisgaardfrantzen H, Bjornvad M, Pedersen S, Schulein M;
 DR WPI; 98-251283/22.
 DR N-PSDB; V29663.
 PT Liquefaction of starch for, e.g. production of sweeteners - comprises use of enzyme hybrids including cellulose binding domain for starch

PS Example 3; Pages 51-53; 83pp; English.
 CC This is the cloned Bacillus agaradherens endoglucanase sequence. This is used in the construction of enzyme hybrids for liquefaction of starch.
 CC The enzyme hybrids contain amino acid sequences of alpha-amylase linked to a cellulose binding domain (CBD). The starch is liquefied by treating in aqueous medium, with such an enzyme hybrid. A recombinant expression vector comprising a construct containing isolated DNA encoding enzyme hybrids with amylolytic activity, promoter and stop signals can be used

CC to transform host cells for the production of the recombinant enzyme
 CC hybrids. The enzyme hybrids are useful in industrial starch processing
 CC especially for the production of sweeteners. Hybrid enzymes have altered
 CC affinity for substrate and increased activity, resulting in at least 1 of
 CC reduced calcium ion dependence, reduced formation of Maillard reaction
 CC products and reduced effect of alpha-amylase on subsequent
 CC saccharification.
 SQ Sequence 462 AA;

Query Match 95.0%; Score 3249; DB 32; Length 462;
 Best Local Similarity 95.5%; Pred. No. 2,686-240;
 Matches 446; Conservative 9; Mismatches 7; Indels 5; Gaps 1;
 Db 1 mkkittifvllmtalfsignttaadsvveehqglslngelvngeqqlkgmss 60
 1 MKKITTIFAVLLMTALFSGNTTAADDYVVEEHGQLSISNGELVNERGEQQLKGMS 60
 Db 61 hglqwygqfvnyesmkwlrddwgitvfraamytssggyldpssvkekvkeaveaaidldi 120
 61 HGLQWYGFVNYESMKWLRDDWGITVFRAMYTSSGGYLDPPSVKEKVKETVEAAIDLGI 120
 QY 61 HGLQWYGFVNYESMKWLRDDWGITVFRAMYTSSGGYLDPPSVKEKVKETVEAAIDLGI 120
 Db 121 yviidwhilsdndpnlykeakdffdemselvgdypnviyeianepngsdvtwngikpy 180
 121 YVIIDWHILSDNDPNLYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 180
 QY 121 YVIIDWHILSDNDPNLYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 180
 Db 181 aeepivirndpnnykeakdffdemselvgdypnviyeianepngsdvtwngikpy 240
 181 AEEVIPVRNDPNNYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 240
 QY 181 AEEVIPVRNDPNNYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 240
 Db 241 dvaldgaafvsewtsaatgvgvfldeaqvwidfmdernlswanwslthkdesaal 300
 241 DYALDGAAFVSEWTSAAATGCGVFLDEAQVWIDFMDERNLSWANWSLTHKDESSAAL 300
 QY 241 DYALDGAAFVSEWTSAAATGCGVFLDEAQVWIDFMDERNLSWANWSLTHKDESSAAL 300
 Db 301 mpgantggvteaelspggtfvrekiresasippsdtpspdpgepdp 355
 301 MPGANTGGVTEAELSPSGTFFVREKIREASIPPSDTPSPDGPDPGPDPTPPSPDG 355
 QY 301 MPGANTGGVTEAELSPSGTFFVREKIREASIPPSDTPSPDGPDPGPDPTPPSPDG 355
 Db 356 kypawdpnglytneivhngqlwqakwtgnqepgdpypgwplksdpsgepdp 415
 356 KYPADPNGLYTNELVHNGQLWQAKWTGNQEPGDPYPGWPLKSDPSGEPDPTPPSD 415
 QY 361 EXPANDSNQIYTNELVHNGQLWQAKWTGNQEPGDPYPGWPLKSDPSGEPDPTPPSD 420
 Db 416 pgeypawdpnglytneivhngqlwqakwtgnqepgdpypgwpln 462
 421 PGEYPADSNQIYTNELVHNGQLWQAKWTGNQEPGDPYPGWPLN 467

RESULT 4
 ID P81843 standard; protein; 409 AA.
 AC P81843;
 DT 10-DEC-1990 (first entry)
 DE Sequence of alkaline phosphatase encoded by Bacillus sp. No. N-4
 DE N-4ACMCase gene
 KW Enzyme.
 OS Bacillus sp. No. N-4.
 PN 162263592-2
 PU 24-NOV-1987
 PF 16-MAY-1986: 111928.
 PR 16-MAY-1986: JP-111928.
 PA (RIKA) Rikagaku Kenkyusho.
 PI WPI: 88-004544/01.
 DR N-PSDB: n82302.
 DT Deoxyribonucleic acid sequence coding cellulase -
 PT is obt'd. from naturally occurring matter, and hybridises with the
 PT DNA of N-4ACMCase gene
 PS Disclosure: Fig 3, Page 479; 12pp; Japanese.
 CC The cellulase gene derived from Bacillus sp. No. N-4 is capable of
 CC producing specific alkaline cellulase. The DNA can be obt'd. from natural
 CC sources or by partial synthesis, and can hybridise with DNA of N-4ACMCase
 CC gene.
 SQ Sequence 409 AA;
 Query Match 82.8%; Score 2832; DB 1; Length 409;

Best Local Similarity 94.9%; Pred. No. 1,476-207;
 Matches 388; Conservative 12; Mismatches 5; Indels 4; Gaps 2;
 Db 1 mkkittifvllmtalfsignttaadsvveehqglslngelvngeqqlkgmss 60
 1 MKKITTIFAVLLMTALFSGNTTAADDYVVEEHGQLSISNGELVNERGEQQLKGMS 60
 Db 61 hglqwygqfvnyesmkwlrddwgitvfraamytssggyldpssvkekvkeaveaaidldi 120
 61 HGLQWYGFVNYESMKWLRDDWGITVFRAMYTSSGGYLDPPSVKEKVKETVEAAIDLGI 120
 QY 61 HGLQWYGFVNYESMKWLRDDWGITVFRAMYTSSGGYLDPPSVKEKVKETVEAAIDLGI 120
 Db 121 yviidwhilsdndpnlykeakdffdemselvgdypnviyeianepngsdvtwngikpy 180
 121 YVIIDWHILSDNDPNLYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 180
 QY 121 YVIIDWHILSDNDPNLYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 180
 Db 181 aeepivirndpnnykeakdffdemselvgdypnviyeianepngsdvtwngikpy 240
 181 AEEVIPVRNDPNNYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 240
 QY 181 AEEVIPVRNDPNNYKEAKDFFDEMSELVGDYPNVIYEIANEPNGSDVTWNGIKPY 240
 Db 241 dvaldgaafvsewtsaatgvgvfldeaqvwidfmdernlswanwslthkdesaal 300
 241 DYALDGAAFVSEWTSAAATGCGVFLDEAQVWIDFMDERNLSWANWSLTHKDESSAAL 300
 QY 241 DYALDGAAFVSEWTSAAATGCGVFLDEAQVWIDFMDERNLSWANWSLTHKDESSAAL 300
 Db 301 mpgantggvteaelspggtfvrekiresasippsdtpspdpgepdp 360
 301 MPGANTGGVTEAELSPSGTFFVREKIREASIPPSDTPSPDGPDPGPDPTPPSPDG 360
 QY 301 MPGANTGGVTEAELSPSGTFFVREKIREASIPPSDTPSPDGPDPGPDPTPPSPDG 360
 Db 361 sdpgypawdpnglytneivhngqlwqakwtgnqepgdpypgwpln 409
 357 SDPGYPADSNQIYTNELVHNGQLWQAKWTGNQEPGDPYPGWPLN 405

RESULT 5
 ID W12379 standard; Protein; 411 AA.
 AC W12379;
 DT 17-JUN-1997 (first entry)
 DE P300-CelB fusion construct 2 polypeptide product.
 DE Cellulase; xylanase; alkaline protease; P300-CelB.
 KW Chimeric-Bacillus-licheniformis-ATCC-53926;
 OS Chimeric-Bacillus sp. N4 (ATCC 21833).
 FH Key
 Location/Qualifiers
 FT peptide 1..26
 FT /label= Sig_peptide
 FT /note= "hybrid between P300 (aal-5) and
 FT CelB (aal-26) signal peptides"
 FT protein 27..411
 FT /label= Mat_protein
 FT /note= "mature CelB cellulase"
 PN W09706181-A1.
 PU 20-FEB-1997
 PF 08-AUG-1996: U12545.
 PR 10-AUG-1995: US-002106.
 PR 07-AUG-1996: US-694346.
 PA (HENK) HENKEL CORP.
 PI Christianson T, Maurer K, Tang MR, Weiss A, Wilson CR;
 DR WPI: 97-154208/14.
 DR N-PSDB: T63227.
 PT System for increased expression of cellulase and xylanase in
 PT Bacillus - Contains gene under control of elements from B.
 PT licheniformis-alkaline protease gene
 PS Disclosure: Fig 8, 8A, 8B; 37pp; English.
 CC The polypeptide product (W12379) of P300-CelB fusion construct 2
 CC (T63227) comprises a hybrid signal peptide, formed between the
 CC signal sequences of Bacillus licheniformis ATCC 53926 (P300)
 CC alkaline protease and Bacillus sp. N4 CelB cellulase, and the
 CC mature CelB alkalophilic cellulase. Expression in Bacillus sp.
 CC host cells of P300-CelB fusion constructs 1-4 (see also W12378,
 CC W12380-81) provides a 10-40 fold improvement of prodn. of the
 CC alkalophilic cellulase in comparison to expression of the native
 CC gene, and an earlier start to the prodn. of enzyme in the
 CC fermentation process.
 SQ Sequence 411 AA;

QY 1 MKKITTIFAVLLMTLALFSIGNTTAADDYSVVEEHGOLSISNGELVNERGEQVQLKGMSS 60
Db 61 hglwqygqfnyesmkwlrddwginvfraamytssggylidpsvkekveaveaaidldi 120
QY 61 HGLWYQGVNYESMKWLRDDWGITVFRAMYTSSGGYIDDPSPKVKETVEAAIDLGI 120
Db 121 yviidwhilsdndpnlykeakdffdemseelydypnvlyeianepngsdvtwgnqikpy 180
QY 121 YVIIDWHILSDNDPNLYKEAKDFFDEMSELYGDYPNVLYEIANEPNGSDVTWGNQIKPY 180
Db 181 aeepvipirndpnllivgtgtsqdvvhhaadnqladpnmvmafhyagthgqnldqv 240
QY 181 AEVIPIRNDPNLLIVGTGTSQDVVHHAADNQLADPNMVMAFHYAGTHGQNLRDQV 240
Db 241 dvaldgaalfvsewgttsaatgvgvfldeaqvwidfmdernlswanwslthkdessaal 300
QY 241 DYALDGAALFVSEWGTSAATGCVGFLDEAQVWIDFMDERNLSWANWSLTHKDESSAAL 300
Db 301 mpganptggwteaelspsgtfvrekiresasippsdtpspdpgepdp-----tppsdpq 355
QY 301 MPGANPTGGWTEAELSPSGTFVREKIREASIPPSDTPSPDPGEPPDPTPPSPDPG 360
Db 356 eypawdpnclytneivyhngqlwqakwvtqngqgdpygpwepln 400
QY 361 EYPAWDSNQIYTNIEIVYHNGQLWQAKWWTQNPQEPGDPYGPWEPLK 405

RESULT 8
ID W22521 standard; Protein; 400 AA.
AC W22521;
DE 08-FEB-1998 (first entry)
DT Bacillus agaradherens alkaline cellulase Cel5A;
KW Desizing; Cellulose; fabric; enzyme hybrid; alkaline cellulase;
KW endoglucanase; Bacillus agaradherens; cellulose binding domain;
KW Cel5A.
OS Bacillus agaradherens strain NCIMB 40482.
PN W09728256-A1.
PD 07-AUG-1997.
PF 29-JAN-1997; DK0041.
PR 29-JAN-1996; DK-000093.
PA (NOVO) NOVO-NORDISK AS.
PI Bjornvad ME, Rasmussen MD, Vind J, Von Der Osten C;
DR WPI; 97-403610/37.
PS N-PSDB; T77055.
PT Desizing cellulose-containing fabric or textile using an enzyme
hybrid - which comprises a catalytically active amino acid sequence
of a non-cellulolytic enzyme linked to an amino acid sequence
comprising a cellulose binding domain.
PT Example 2; Page 52-54; 72pp; English.
PS This protein comprises the alkaline cellulase (endoglucanase) of
Bacillus agaradherens NCIMB 40482. It was expressed in Bacillus
subtilis P2306 transformants following PCR amplification (see
T77053-54) of B. agaradherens genomic DNA and ligation of the PCR
product into vector pDN1981. DNA encoding the cellulose binding
domain (CBD) of the alkaline cellulase was subsequently amplified
(see T77056-57) for use in the construction of a novel alpha-
amylase-CBD hybrid enzyme (see T77058). A claimed process for
desizing cellulose-containing fabric or textile comprises treating
the fabric or textile with a modified enzyme (enzyme hybrid)
comprising a non-cellulolytic enzyme linked to a CBD. The process
gives improved enzyme performance by modifying the enzyme so as to
increase its affinity for cellulosic fabric.
SQ Sequence 400 AA;

Query Match 81.3%; Score 2781; DB 25; Length 400;
Best Local Similarity 95.6%; Pred. No. 1.47e-203;
Matches 387; Conservative 8; Mismatches 5; Indels 5; Gaps 1;
Db 1 mkkittifavllmtlalfsignttaaddysvveehgolsisngelvnrggeqvlkgmss 60
QY 1 MKKITTIFAVLLMTLALFSIGNTTAADDYSVVEEHGOLSISNGELVNERGEQVQLKGMSS 60
Db 61 hglwqygqfnyesmkwlrddwginvfraamytssggylidpsvkekveaveaaidldi 120

QY 61 HGLWYQGVNYESMKWLRDDWGITVFRAMYTSSGGYIDDPSPKVKETVEAAIDLGI 120
Db 121 yviidwhilsdndpnlykeakdffdemseelydypnvlyeianepngsdvtwgnqikpy 180
QY 121 YVIIDWHILSDNDPNLYKEAKDFFDEMSELYGDYPNVLYEIANEPNGSDVTWGNQIKPY 180
Db 181 aeepvipirndpnllivgtgtsqdvvhhaadnqladpnmvmafhyagthgqnldqv 240
QY 181 AEVIPIRNDPNLLIVGTGTSQDVVHHAADNQLADPNMVMAFHYAGTHGQNLRDQV 240
Db 241 dvaldgaalfvsewgttsaatgvgvfldeaqvwidfmdernlswanwslthkdessaal 300
QY 241 DYALDGAALFVSEWGTSAATGCVGFLDEAQVWIDFMDERNLSWANWSLTHKDESSAAL 300
Db 301 mpganptggwteaelspsgtfvrekiresasippsdtpspdpgepdp-----tppsdpq 355
QY 301 MPGANPTGGWTEAELSPSGTFVREKIREASIPPSDTPSPDPGEPPDPTPPSPDPG 360
Db 356 eypawdpnclytneivyhngqlwqakwvtqngqgdpygpwepln 400
QY 361 EYPAWDSNQIYTNIEIVYHNGQLWQAKWWTQNPQEPGDPYGPWEPLK 405

RESULT 9
ID W12378 standard; Protein; 410 AA.
AC W12378;
DE 17-JUN-1997 (first entry)
DT P300-CelB fusion construct 1 polypeptide product.
KW Cellulase; xylanase; alkaline protease; P300; CelB.
OS Chimeric Bacillus licheniformis ATCC 53926;
OS Chimeric Bacillus sp. N4 (ATCC 21833).
FH Key Location/Qualifiers
FT peptide 1..26
FT /label= Sig_peptide
FT /note= "hybrid between P300 (aal-12) and
CelB (aal3-26) signal peptides"
FT protein 27..410
FT /label= Mat_protein
FT /note= "mature CelB cellulase"
PN W09706181-A1.
PD 20-FEB-1997.
PF 08-AUG-1996; UI2545.
PR 10-AUG-1995; US-002106.
PR 07-AUG-1996; US-694346.
PA (HENK) HENKEL CORP.
PI Christianson T, Maurer K, Tang MR, Weiss A, Wilson CR;
DR WPI; 97-154208/14.
DR N-PSDB; T63226.
PT System for increased expression of cellulase and xylanase in
Bacillus - contains gene under control of elements from B.
PT licheniformis alkaline protease gene
PS Disclosure; Fig 7,7A,7B; 37pp; English.
CC The polypeptide product (W12378) of P300-CelB fusion construct 1
CC (T63226) comprises a hybrid signal peptide, formed between the
CC signal sequences of Bacillus licheniformis ATCC 53926 (P300)
CC alkaline protease and Bacillus sp. N4 CelB cellulase, and the
CC mature CelB alkalophilic cellulase. Expression in Bacillus sp.
CC host cells of P300-CelB fusion constructs 1-4 (see also
CC W12379-81) provides a 10-40 fold improvement of prodn. of the
CC alkalophilic cellulase in comparison to expression of the native
CC gene, and an earlier start to the prodn. of enzyme in the
CC fermentation process.
SQ Sequence 410 AA;

Query Match 81.1%; Score 2772; DB 21; Length 410;
Best Local Similarity 92.9%; Pred. No. 7.46e-203;
Matches 380; Conservative 17; Mismatches 8; Indels 4; Gaps 2;
Db 2 mrksfvlgmllmtlalfsignttaaddysvveehgolsisngelvnrggeqvlkgmss 61
QY 1 MKKITTIFAVLLMTLALFSIGNTTAADDYSVVEEHGOLSISNGELVNERGEQVQLKGMSS 60

QY	1	MKKTTTTFVALLMTLAL-FSIGN-TTAAADDYSVVEEHGOLSISSNGELVNERGEQVQLKGM	58
Db	61	sshgqlwygcfvnyesmkwlrddgwtvtfraamytssggyiedpsvkekveaveaaidl	120
QY	59	SSHGLQWGGFVWYESMKWLRDDGWTVTFRAAMTSSGGTIDDPVSVEKVKETVEAIDL	118
Db	121	giyviidwhilsdndpnikeeakdfdemselgydpnviyeianepngsdvtdwnqik	180
QY	119	GIVYIIDWHILSDNDPNYKEEAKDFDEMSELYGDPNVVIEIANEPNGSDVTDWNQIK	178
Db	181	pyaevipvrindpnlniivtgtwtswdvhhaadnqltdpnvmvafhyagthgqnldr	240
QY	179	PYAEVIPVRINDPNVNIIVTGTWSDVHHAADNQLADPNVMVAFHFYAGTHGQNLRD	238
Db	241	qvdyalddgaafvsewgtseatgqgvfideaqvwlfdmderlniswanwsithkdesa	300
QY	239	QVDYALDDGAAFVSEWGTSAATGCGVFLDEAQVWIDFMDERNISWANWSITHKDESSA	298
Db	301	almpgasptgwtweaelspgstfvrekiresattpssdptpsdpdpgepdpgepdppt	360
QY	299	ALMPGANPTGWTWEAELSPGSTFVREKIREASIPSPDPTPFSDP--GEP--DPGEPT	354
Db	361	ppsdpgdypawdntlytdeivhngqlwgakwvtanqegpgdypgwpeln	411
QY	355	PPSDPGEYPAWDSNQIYTNIEIVHNGQLWQAKWWTQNPQPGDYPGWPPEPLK	405
RESULT 11			
ID	W12380	standard; Protein; 412 AA.	
AC	W12380;		
DT	17-JUN-1997	(first entry)	
DE	P300-CelB fusion construct 3	polypeptide product.	
KW	Cellulase; xylanase; alkaline protease;	p300; CelB.	
OS	Chimeric Bacillus licheniformis	ATCC 53926;	
OS	Chimeric Bacillus sp. N4	(ATCC 21833).	
FH	Key	Location/Qualifiers	
FT	FT	1..29	
FT	FT	/label= Sig_peptide	
FT	FT	/note= P300 signal peptide"	
FT	FT	30..412	
FT	FT	/label= Mat_protein	
FT	FT	/note= "mature CelB cellulase"	
FT	FT	WO9706181-AL.	
PN	PD	20-FEB-1997.	
PF	PF	08-AUG-1996; U12545.	
PR	PR	10-AUG-1995; US-002106.	
PR	PR	07-AUG-1996; US-694346.	
PA	PA	(HENK) HENKEL CORP.	
PI	PI	Christianson T, Maurer K, Tang MR, Weiss A, Willson CR;	
DR	DR	WPI; 97-154208/14.	
DR	DR	N-PSDB; T63328.	
PT	PT	System for increased expression of cellulase and xylanase in	
PT	PT	Bacillus - contains gene under control of elements from B.	
PS	PS	licheniformis alkaline protease gene	
PS	PS	Disclosure; Fig 9.9A,9B; 37pp; English.	
CC	CC	The polypeptide product (W12380) of P300-CelB fusion construct 3	
CC	CC	(T63328) comprises the signal peptide of Bacillus licheniformis	
CC	CC	ATCC 53926 (P300) alkaline protease and the mature CelB	
CC	CC	alkalophilic cellulase of Bacillus sp. N4. Expression in Bacillus	
CC	CC	sp. host cells of P300-CelB fusion constructs 1-4 (see also W12378-	
CC	CC	79, W12381) provides a 10-40 fold improvement of prodn. of the	
CC	CC	alkalophilic cellulase in comparison to expression of the native	
CC	CC	gene, and an earlier start to the prodn. of enzyme in the	
CC	CC	fermentation process.	
SQ	SQ	Sequence 412 AA;	
Query Match 78.7%; Score 2691; DB 21; Length 412;			
Best Local Similarity 90.3%; Pred. No. 1 68e-196;			
Matches 364; Conservative 28; Mismatches 6; Indels 5; Gaps			
Db	11	mltafm1vftm-afdsasaddyvveehqqlsisngeivndrgepvqlgmshglqwy	69
QY	7	IFVALLMTLALFIGNTTAADDYSVVEEHGOLSISSNGELVNERGEQVQLKGMSSHGLQWY	66

Db 70 gqfvnyesmkwlrddwgitvfraamytssgyiedpsvkekveaeadlgiyyidw 129
 QY 67 GQFVNYESMKWLRDDWGITVFRAMYTSSGYIDPVSKEKVEAEADLGIYVIDW 126
 Db 130 hilsdnpnylkeakdffdemselvdygnvviyeianepngsdvtdnqikpyaevip 189
 QY 127 HILSDNDPNYIKEAKDFFDEMSELYGDPNVIYEIANEPNGSDVTWQNKPYAEVIP 186
 Db 190 virndpnniivgtwsgdvhaadnqltdpnmvafhyagthgnlrdqdyaldq 249
 QY 187 VIRNDPNNIVGTWSDVHAADNQLADPNVAFHYAGTHGNLRDQDYALDQ 246
 Db 250 gaafvsewtsatgsgvfldeaqvdfmdernlswanwlthkdessaaampgasp 309
 QY 247 GAAIFVSEWTSATGSGVFLDEAQVDFMDERNLSWANWLTHTKDESSAALMPGANP 306
 Db 310 tggwteaelspgftvrekiresattppsdtpptpsdpgepdpptpsdpgdy 369
 QY 307 TGGWTEAELSPGFTVREKIRESATTPPSDTPPTSPDGEPPPTSPDGEY 362
 Db 370 pawdntiyltdelvyhngqlwqakwtqngqpgdpygpwepln 412
 QY 363 PAWDSNQIYNEIVYHNGQLWQAKWTQNGQPGDYPGPWEPLK 405

RESULT 12
 ID R42122 standard; Protein; 499 AA.
 AC R42122;
 DT 27-APR-1994 (first entry)
 DE NK-1 cellulase.
 KW Cellulase; pH dependence; mutation.
 OS Bacillus N4.
 PN J05236969-A.
 PD 17-SEP-1993.
 PF 28-FEB-1992; 075883.
 PR 28-FEB-1992; JP-075883.
 PA (BEPF) BEPPU T.
 DR WPI: 93-330585/42.
 DT N-PSDB; Q49820.
 PT Changing the pH-dependence of cellulase enzymatic activity - by
 PT changing base sequence of cellulose-producing gene of Bacillus
 PT microbe to base sequence coding asparagine and serine residues at
 PT specified aminoacid sites
 PS Disclosure: Fig 1: 9pp; Japanese.
 CC Sequences (Q49818-19) consist of two synthetic fragments which are
 CC used to induce a mutation within the cellulase NK-1 gene. The gene
 CC is shown in sequence (Q49820). The modified NK-1 gene shows a
 CC change in pH dependence.
 SQ Sequence 499 AA;

Query Match 48.2%; Score 1649; DB 8; Length 499;
 Best Local Similarity 62.7%; Pred. No. 4.80e-115;
 Matches 222; Conservative 62; Mismatches 64; Indels 6; Gaps 6;
 Db 1 mkrsisifitcllitvmtgglqspasaaqtktpaaknglsikgtqlvnrdrkavqlk 60
 QY 1 MKKITTFIVAF-LMT-LALFSI-GNTTAAADYSV-VEEHGQLSINSGELVNERGEQVLK 56
 Db 61 gishgqlwqvgdvfnkdsllkwrddwgitvfraamytdaggyidnpsvknkvkeaveak 120
 QY 57 GMSHGLQWYQGFVNYESMKWLRDWDGITVFRAMYTSSGGYIDPVSKEKVEAEVAAI 116
 Db 121 elgilyiidwhilndgpnqhkakaffkemsslyngntpnviyeianepng-dvnrkrd 179
 QY 117 DLGIYVIDWHILSDNDPNYIKEAKDFFDEMSELYGDPNVIYEIANEPNGSDVTWQNK 176
 Db 180 ikpyaevisvirkndpndnliivgtgtwsgdvdaaddqlkdanvmyalhfyagthgsl 239
 QY 177 IKPYAEVIVIRNDPNNIVGTGTWSDVDAADDQLKDNVMYALHFYAGTHGQNL 236
 Db 240 rdkanyalskgaplfvtegwtsdagngvfldegrewlnyldsknlswnwnlsdkges 299

QY 237 RDQVDYALDQGAALFVSEWTSATGSGVFLDEAQVDFMDERNLSWANWLTHTKDES 296
 Db 300 ssalkpgasktgwpltdltasgtfvrenilngkndstkerpetpaqdnpageng 353
 QY 297 SAALMPGANPTGGWTEAELSPGFTVREKIRESATIPSPDP-TPSPDGEPPDPG 349
 RESULT 13
 ID W18790 standard; protein; 551 AA.
 AC W18790;
 DT 18-NOV-1997 (first entry)
 DE Corrected Bacillus lautus (NCIMB 40250) endoglucanase Endo 3A.
 KW Endoglucanase; Endo 3A; formation; localised; variation;
 KW colour density; surface; dye; fabric; family 5; cellulose;
 KW hydrolysatn; p-nitrophenyl-beta-1,4-cellobioside; stone wash;
 KW blue jeans; back staining.
 OS Bacillus lautus.
 PN W09709410-A1.
 PD 13-MAR-1997.
 PF 03-SEP-1996; DK0364.
 PR 08-SEP-1995; DK-000993.
 PA (NOVO) NOVO-NORDISK AS.
 PI Fich M, Onishi M, Schulein M, Toft AH;
 DR WPI: 97-192888/17.
 PT Localised variation of colour density in the surface of a dyed
 PT cellulosic fabric - uses cellulase compsn. able to hydrolyse
 PT p-nitrophenyl-beta-1,4-cellobioside
 PS Disclosure: Pages 15-17; 23pp; English.
 CC The present sequence is the corrected version of the incorrect
 CC Bacillus lautus (NCIMB 40250) endoglucanase Endo 3A described in
 CC W09110732. Endo 3 can be used in novel method of forming localised
 CC colour density variation on the surface of a dyed cellulosic
 CC fabric. The method comprises agitating the fabric in an aqueous
 CC medium (pH 6.5 to 9.0) containing a family 5 cellulose,
 CC e.g. the present sequence, which can hydrolyse p-nitrophenyl-beta
 CC -1,4-cellobioside, or a family 7 cellulase, and a mechanical
 CC abrading agent or cellulose having abrading activity. Each
 CC cellulase displays 30 % or more of its maximum activity at pH 7.
 CC The process is useful to provide a stone washed look to blue jeans
 CC without back staining.
 SQ Sequence 551 AA;

Query Match 45.7%; Score 1561; DB 24; Length 551;
 Best Local Similarity 62.3%; Pred. No. 3.38e-108;
 Matches 192; Conservative 63; Mismatches 51; Indels 2; Gaps 2;
 Db 7 gqkvqngqlvgsgqavqlvgmshglqwygnfnvkslqwmrdnwglnvfraamytae 66
 QY 36 GQLSINGELVNERGEQVLKGMSSHGLQWYGFVNYESMKWLRDWDGITVFRAMYTSS 95
 Db 67 dgyitdpsvknkvkeavqasidlglyviidwhilsdgnpntykaqskaffgematlygnt 126
 QY 96 GGYIDDPVSKEKVEAEADLGIYVIDWHILSDNDPNYIKEAKDFFDEMSELYGDPY 155
 Db 127 pnviyeianepng-nvswad-vksyaeveitairaidpdpvviygsptwsqdihaadnp 184
 QY 156 PNVIYEIANEPNGSDVTWQNKPYAEVIVIRNDPNNIVGTGTWSDVHAADNQ 215
 Db 185 vshsnvmalfysgthqgfirdrityamkgaalfvtegwtsdagnggpyfpqskewi 244
 QY 216 LADPNVMTAFHYAGTHGNLRDQDYALDQGAALFVSEWTSATGSGVFLDEAQVWI 275
 Db 245 dflnarkiswnwnlsadkvetssalmpgasptggvtdaqlesgkwrdqirqtggsg 304
 QY 276 DFMERNLSWANWLTHTKDESSAALMPGANPTGGWTEAELSPGFTVREKIRESATIPSP 335
 Db 305 nptapaap 312
 QY 336 DPTPPSDP 343

RESULT 14
 ID W01503 standard; protein; 531 AA.

AC W01503;
 AT 26-FEB-1997 (first entry)
 DE 60 kb endoglucanase, EG C.
 KW Detergent composition; cellulase; retaining-type activity; catalytic;
 KW activity; cellotriose; particulate soil removal; colour clarification;
 KW cleaning; cellulose-containing fabric; cellubiohydrolase; endoglucanase.
 OS Bacillus lautus, NCIMB 40250.
 PN W09502675-A1.
 PD W09502675-A1.
 PF 26-JAN-1999;
 PF 07-JUL-1994; DK0280.
 PR 12-JUL-1993; EP-870131.
 PR 11-OCT-1993; DK-001135.
 PA (NOVO) NOVO-NORDISK AS.
 PA (PROC) PROCTER & GAMBLE CO.
 PI Convents AC, Jeffreys B, Schuelein M, Tikhomirov DF;
 PI WPI; 95-067325/09.
 PT Detergent compsn. contg. two cellulase components - the first
 PT removing soil particles and the second capable of colour
 PT clarification, useful in laundry compsns.
 PS Claim 28; Page 68-70; 83pp; English.
 CC Detergent compositions comprising: 1) a first cellulase component having
 CC retaining-type activity, pref. having a catalytic activity on cellotriose
 CC at pH 8.5 corresponding to that of at least 0.01 s<-1>, and capable of
 CC particulate soil removal; and 2) a second cellulase component having
 CC multiple domains comprising at least one non-catalytic domain attached to
 CC a catalytic domain, pref. having a catalytic activity on Red Avicel 7.5
 CC per 1 mg of cellulase protein higher than 10<-4> IU and being capable of
 CC colour clarification, where at least one of the cellulase components is a
 CC single (recombinant) component, are useful for cleaning and colour
 CC clarification of cellulose-containing fabrics. The second cellulase
 CC component can be an endoglucanase which is immunoreactive with an
 CC antibody raised against a highly purified ~60 kD endoglucanase
 CC derived from Bacillus lautus, NCIMB 40250, and is esp. the present
 CC sequence, designated EG C.
 SQ Sequence 531 AA;

Query Match 22-28; Score 1339; DB 20; Length 531;
 Best Local Similarity 58.5%; Pred. No. 5.84e-91;
 Matches 176; Conservative 52; Mismatches 68; Indels 5; Gaps 4;
 Db 13 vltmvlmgl1llpygarkgaa-pavp-fgelkvqngqlvgsgqavqlvgmsshglqwy 72
 QY 10 VLMTLALFSGNTAADDYSVVEEHGQLSISNGELVNERGEQVQLKGMSSHGLQWY 68
 Db 73 fvnksslqwmrdnwginvfraamytsgdyitdpsvknkveavqasidlalyviidwhi 132
 QY 69 FVNYESMKWLRDDMGITVFRAMYTSSGGYIDDPVKEKVEAEADLGIYVIIDWHI 128
 Db 133 lsdgnpntykaqskaffqematlygnpviyeiarstec-vl--grcqs-seevital 188
 QY 129 LSONDPNIYKEEAKDFDEMSLEYGDYPNVIYEIANEPNGSDVTWQDNIKPYAEVPIV 188
 Db 189 rtsldpdrvvgvsgptsqdihlaadnpvshnvmvalhfygthgqfdrityamnkga 248
 QY 189 RNDPNPNIYVIGTGSQDVHHAADNPVNYAFHFYAGTHGQNLRDQVDYALDQGA 248
 Db 249 aifvtewgtsdasnggpylpqskewidflnarkiswvnwsladkvettsaalmpgasp 308
 QY 249 AIFVSEWGTSAATGCGVFLDEAOWWIDFDERNLNLSWANNSLTHKDESSAALMPCANFTG 308
 Db 309 a 309
 QY 309 G 309

RESULT 15
 ID R13229 standard; Protein; 532 AA.
 AC R13229;
 DT 14-OCT-1991 (first entry)
 DE Endoglucanase encoded by endo3 gene.
 KW Cellulase activity; detergent.
 OS Bacillus spp. NCIMB 40250.
 FH Key Location/Qualifiers

FT cleavage_site 36..37
 FT peptide 1..36
 FT peptide /label= signal peptide
 FT peptide 37..531
 FT peptide /label= mature peptide
 PN W09110732-A.
 PD 25-JUL-1991.
 PF 18-JAN-1991; DK0013.
 PR 19-JAN-1990; DK-000164.
 PA (NOVO) NOVO NORDISK A/S.
 PI Jorgensen PL, Schulein M, Hansen C;
 DR WPI; 91-238020/32.
 DR N-PSDB; Q13003.
 PT Enzyme exhibiting cellulase activity from Bacillus sp. - is an
 PT endo-glucanase, esp. useful for harshness redn. of cotton-contg.
 PT fabrics.
 PS Claim 1; Page 80; 96pp; English.
 CC The enzyme is encoded by a 11000 bp. EcoRI fragment of Bacillus
 CC spp. PL236 DNA contained in plasmid pPL591. It exhibits an
 CC endoglucanase activity of at least 10 (pref. at least 25)
 CC carboxymethyl cellulose (CMC) endoase units per mg total protein
 CC under alkaline conditions. It is especially useful as a
 CC cellulolytic agent and has been found to be more stable during
 CC washing (60 mins. at 40 deg.) in the presence of conventional
 CC detergents than a commercial cellulase preparation. It may also
 CC show increased storage stability in liq. detergents contg.
 CC proteases. The sequence was deduced from the DNA (Q13003), it is a
 CC product of the endo3 gene. See also R13227 and R13228.
 SQ Sequence 532 AA;
 Query Match 38.9%; Score 1329; DB 3; Length 532;
 Best Local Similarity 56.8%; Pred. No. 3.48e-90;
 Matches 172; Conservative 61; Mismatches 64; Indels 6; Gaps 5;
 Db 13 vltmvlmgl1llpygarkgaa-pavp-fgelkvqngqlvgsgqavqlvgmsshglqwy 70
 QY 7 fAVLMTALFSGNTAADDYSVVEEHGQLSISNGELVNERGEQVQLKGMSSHGLQWY 66
 Db 71 gnfvnksslqwmrdnwginvfraamytsgdyitdpsvknkveavqasmdlalyviidw 130
 QY 67 GQFVNYESMKWLRDDMGITVFRAMYTSSGGYIDDPVKEKVEAEADLGIYVIIDW 126
 Db 131 hilsdgnpntykaqskaffqematlygnpviyeiarstec-vl--grcqs-seevit 186
 QY 127 HTLSDNDPNIYKEEAKDFDEMSLEYGDYPNVIYEIANEPNGSDVTWQDNIKPYAEVPI 186
 Db 187 airtsldpdrvvgvsgptsqdihlaadnpvshnvmvalhfygthgqfdrityamnk 246
 QY 187 VIRNDPNPNIYVIGTGSQDVHHAADNPVNYAFHFYAGTHGQNLRDQVDYALDQ 246
 Db 247 gaaifvtewgtsdasnggpylpqskewidflnarkiswvnwsladkvettsaalmpgasp 306
 QY 247 GAAIFVSEWGTSAATGCGVFLDEAOWWIDFDERNLNLSWANNSLTHKDESSAALMPCANP 306
 Db 307 tga 309
 QY 307 TGG 309

Search completed: Fri Jun 4 10:00:43 1999
 Job time : 82 secs.

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	604	14.3	414	2 Q59733	BETA-1,4-D-GLUCANASE (1.25e-97
2	567	13.4	477	3 Q13333	CELLULOSE.	1.08e-89
3	535	12.6	400	2 Q08342	CELLULOSE (EC 3.2.1.4)	7.24e-83
4	535	12.6	1232	3 Q59943	CELLULOSE CELD (FRAGME	7.24e-83
5	528	12.5	482	3 Q01409	CELLULOSE (FRAGMENT).	2.23e-81
6	517	12.2	471	3 P78719	CELLULOSE.	4.81e-79
7	513	12.1	388	3 Q13334	CELLULOSE.	3.38e-78
8	470	11.1	455	2 Q53302	CELLULOSE.	3.92e-69
9	465	11.0	759	2 Q05143	CELLULOSE.	4.38e-68
10	447	10.6	363	2 Q04878	CARBOXYMETHYLCELLULOSE	2.53e-64
11	447	10.6	924	2 Q06842	B14 PUTATIVE POLYGALA	2.53e-64
12	396	9.4	519	2 Q47916	ENDOGLUCANASE CELG.	9.06e-54
13	304	7.2	584	2 Q45397	XYLANASE PRECURSOR.	2.99e-35
14	156	3.7	490	2 Q33853	ENDOGLYCERAMIDASE II	3.51e-08
15	141	3.3	419	3 Q12539	EXO-1,3-BETA-GLUCANASE	8.25e-06
16	137	3.2	410	3 Q12665	ENDOGLUCANASE 2 (EC 3.	3.38e-05
17	130	3.1	416	3 Q01257	43 KDA SECRETED GLYCOP	3.79e-04
18	126	3.0	368	3 Q12637	ENDO-BETA-1,7-GLUCANAS	1.46e-03
19	123	2.9	494	2 Q86099	CELLULOSE PRECURSOR (E	3.94e-03
20	119	2.8	160	3 Q12540	EXO-1,3-BETA-GLUCANASE	1.45e-02

Db 304 NADWELFEWDNTPATAELITLMSNLKENYLDKDIPVIITEYG 345

```

QY      |  || ::  |:|  :  ::  |||::  |:|
261 -AGYTRFE-EDS--KREIITFDRVHHTFVARGIPVVLGEFG 298

```

```

RESULT      2
ID          013333
AC          PRELIMINARY;
AD          PRT; 477 AA.
DT          01-JAN-1998 (TREMBLREL. 05, CREATED)
DT          01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT          01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE          CELLULOSE.
DE          CELB29.
GN          ORPINOMYCES JOYONII.
OS          EUKARYOTA; FUNGI; CHYTRIDIOMYCOTA; CHYTRIDIOMYCETES;
OC          NEOCALLIMASTICALES; NEOCALLIMASTICACEAE; ORPINOMYCES.
[1]
RN          SEQUENCE FROM N.A.
RC          STRAIN-SG4:
RA          QIU X., SELINGER L.B., YANKE L.J., CHENG K.-J.;
RL          SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR          EMBL; AF015248; G2353005; -.
DR          PFAM; PF00150; cellulase; 1.
SQ          SEQUENCE 477 AA; 53664 MW; 38088821C CRC32:

```

Query Match	13.4%;	Score 567;	DB 3;	Length 477;
Best Local Similarity	29.9%;	Pred. No. 1.08e-99;		
Matches	103;	Conservative	88;	Mismatches 136;
			Indels	18;
			Gaps	13;

Db	1	MKFFKNTLALLTPVIAGSNAWRNTPS-KDLVKELNIGNLGNALDAHCLDKLDKDYNDQLA	59
Qy	1	MKWK-SWVLAVLVVSVFVAPVSSANEDKTLDIQSVYRDMQPGNULGN-TFDAVG-Q	57
Db	60	SETCAWPAKPTGFLSALKNGQNVFPIPTWTGFGNGDPDYKISDVMMRRVHEVVDYAL	119
Qy	58	DETAGNPRVTRELIERIADGEYKSIRIPWTENRIGGADPYDIPQFLNRVDEVVQWAL	117
Db	120	NTGGSVILNTHHEW--NYAFSNNLQKAKPILAAIKWQIAAEFANYDEHLIFGGMNEPK	177
Qy	118	EEDLYVMIHLHDSWLVIYEMEHYNGVMARYSLWEQLSNHFKDPTKLMFSVNEP-K	176
Db	178	VDHPEWNGGDKGWDFVNEWNAVFLQTVRASGNNAIRHLMTPTAACVNNGALSESYPK	237
Qy	177	FSQ-N-WGETRENHHAALLDQNTVFVIVRQSGQNDIRPLVPTMETATSQPLLNNLYQ	234
Db	238	KSPT-NDNKVIASVHSYVPYNFALNTCAGAEK-TFGSTSDIEAMNNIKRFLVDRIPIVI	295
Qy	235	TIDKLDDPNLIATVHYGYGFWPFSVNI-AGYTR EEDSKRELIETTFDRVHHTFVARGIPV	293
Db	296	IGEFAM--NRDN----ESERARWAEEYIKSATAMGVPVCLWDNG	334
	294	LGEFGLLGFKQHTGVIQOGEKLFKPFYLIHHLNERDITTHLWDNG	338

RESULT	3	PRELIMINARY;	PRT;	400 AA.
AD	O08342			
AC	O08342;			
DT	01-JUL-1997	(TREMBREL. 04, CREATED)		
DT	01-JUL-1997	(TREMBREL. 04, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBREL. 08, LAST ANNOTATION UPDATE)		
DE	CELLULASE (EC 3.2.1.4)	(ENDOGLUCANASE)		
DE	(CARBOXYMETHYL CELLULASE).			
GN	CELA.			
OS	BACILLUS SP.			
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;			
OC	BACILLUS.			
[1]				
RN	SEQUENCE FROM N.A.			
RP				
RC	STRAIN=BP-23:			
RA				
RL	BLANCO A., PASTOR F. I. J. ;			
RL	SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.			
CC	-1- CATALYTIC ACTIVITY; ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC			
CG	*LINKAGES IN CELLULOSE.			
DR	ENBL; Y12512; E311859; -			

[illegible]

```

RESULT      4
ID          OS9943
AC          OS9943;
DT          01-AUG-1998 (TREMBLEL. 07, CREATED)
DT          01-AUG-1998 (TREMBLEL. 07, LAST SEQUENCE UPDATE)
DT          01-NOV-1998 (TREMBLEL. 08, LAST ANNOTATION UPDATE)
DE          CELLULOSE CELD (FRAGMENT).
GN          CELD.
OS          NEOCALLIMASTIX PATRICIARUM (RUMEN FUNGUS).
OC          EUKARYOTA; FUNGI; CHYTRIDIOMYCOTA; CHYTRIDIOMYCETES;
NC          NEOCALLIMASTICALES; NEOCALLIMASTICACEAE; NEOCALLIMASTIX.
RN          [1]
RP          SEQUENCE FROM N.A.
RX          MEDLINE; 93123992.
RX          XUE G.P., GOBIUS K.S., ORPIN C.G.;
RT          "A novel polysaccharide hydrolase cDNA (celd) from Neocallimastix patriciarum encoding three multi-functional catalytic domains with high endoglucanase, cellobiohydrolase and xylanase activities."
RL          J. GEN. MICROBIOL. 138:0-0(0).
DR          EMBL; AF053363; G2981484; -.
DR          NC_001111
SQ          SEQUENCE 1232 AA; 140617 MW; E4C60A2B CRC32;

```

Query Match	12.6%	Score 535;	DB 3;	Length 1232;
Best Local Similarity	31.1%	Pred. No. 7.24e-83;		
Matches	89;	Conservative 81;	Mismatches 98;	Indels 18; Gaps 15;
Db	828	ETCWGNPKTTEDMFKVLMDNQFNVPRTPTWSGHFGSEADPYKINEKWLKRHVEIVDPYK	897	
QY	59	ETAWGNPRVREIERIADEGYKSIRIPVTWENRIGAPDYPIDPOFLNRVDVVGWALE	118	
Db	888	NGAFVIILNLHETWNHAFS-E-TLDTAKELLEIKWSQIAEEFKDYDEHLIFEGELNPKRN	945	
QY	119	EDLVFMNLNHDLSWLTWYEMEHYNGVMKYRSLWFQSLSNHFKDYPTKLMEFVSNPEP-KF	177	
Db	946	DTVPVEWTGGQDEGWAV-NAMNAVFLKTIRSSGNNPKRHLMPYPAAACNENSFKN-F-	1002	
QY	178	-SQ-NW-GEIRENNHALLDLNLVFEIVRQSGQNDIRPLVPTMETATSQPLNNLYQ	234	
Db	1003	IFPE-DDDKVIASVHAYAFYNFALNRNGAGAVDKFDAAGKDKLEWNINMLMKRREVDOGIPM	1061	
QY	235	TIDKLDDPNLIATVHYVYGFWPFSVNI-AGYT-REFEEDSKREIETDFRVHHTFEVARGIPV	292	
Db	1062	ILGEYGM--NRDN---EE-ERATWAEFYMEKVTAMGVQVQWWDNG	1101	
QY	293	VLGEFGLLGDFKDTGVIQOQEKLFKEFYTLHNLNERDITHMLWDNG	338	

[illegible]

<hr/>					
RESULT	8	PRELIMINARY;	PRT;	455 AA.	
ID	Q33302	AC	Q33302;		
DT	01-NOV-1996	(TREMBREL, 01, CREATED)			
DT	01-NOV-1996	(TREMBREL, 01, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMBREL, 08, LAST ANNOTATION UPDATE)			
DE	CELLULOSE.				
GN	ENDA.				
OS	RUMINOCOCCUS FLAVEFACIENS.				
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;				
OC	RUMINOCOCCUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 91360084.				
RA	CUNNINGHAM C., MCPHERSON C.A., MARTIN J., HARRIS W.J., FLINT H.J.;				
RT	"Sequence of a cellulase gene from the rumen anaerobe Ruminococcus				
RT	flavefaciens 17."				
RL	MOL. GEN. GENET. 228:320-323(1991).				
DR	EMBL; S55178; G234872; -.				
DR	PFAM; PF00150; cellulase; 1.				
SQ	SEQUENCE 455 AA; 52420 MW; 7C883649 CRC32;				
<hr/>					
	Query Match	11.1%;	Score 470;	DB 2:	Length 455;
	Best Local Similarity	29.4%;	Pred. No. 3.92e-09;		
	Matches	78;	Conservative	68;	Mismatches 102;
				Indels	17;
				Gaps	16;
<hr/>					
Db	74	LDSAGLETETCWGCPASQELFDAIKAKGFNTVRIPITTFWQHLDEN-DN-IDPAMARVH 131	:	: : :	: : : :
Qy	52	FVDVGQD-ETAWGNPRVTRELIERIADEGVKSRIPYTWNRNIGGAPDPIDPQFLNRVD 110	:	: : :	: : : :
Db	132	QVDVAYNIGLYVIINHHEQNWINRADLATAYDDINPLMKLTQTATFEKDYDQHLIF 191	:	: : : :	: : : :
Qy	111	EVOGWALEEDLYMINLHD-SWLWIYEMEHNYNGVMKYRSLWEQLSNHF KDYP TKLMF 169	:	: : : :	: : : :
Db	192	ECKNPERAMTTPPWWSATPVESRDVINLEAFVELIRGMDFYAKTRLLMLPGYVASS 251	:	: : :	: : :
Qy	170	ESVNEPK-F-SQ-NW-GEIRENHALLDDLTFTFFEVRSQGQ-QNDIRPLVLPTMETAT 224	:	: : : :	: : : :
Db	252	DKTFLNQIVLP-EN-DD-FLAVSHIATPYNTMTNKTEBGAYHDTTFKEFSNDLAYNLQ 308	:	: : : : :	: : : : : :
Qy	225	SQPLNNLYQTIDKLDPNLIATVHYGFPPFSVNI-A--G-Y-TFREEDSKREIIETF 279	:	: : : :	: : : :
Db	309	NFRDMFINKDIPIVVIGMGTSDFGN 333	:	: : : :	: : : :
Qy	280	RVHHTFVARGIPVVLGEFGLLGFDK 304	:	: : : :	: : : :
<hr/>					
RESULT	9	PRELIMINARY;	PRT;	759 AA.	
ID	O05143	AC	O05143;		
DT	01-JUL-1997	(TREMBREL, 04, CREATED)			
DT	01-JUL-1997	(TREMBREL, 04, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMBREL, 08, LAST ANNOTATION UPDATE)			
DE	ENDOGALACTANASE A PRECURSOR.				
GN	ENDA.				
OS	RUMINOCOCCUS FLAVEFACIENS.				
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;				
OC	RUMINOCOCCUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=17;				
RC	MEDLINE; 97286515.				
RA	KIRBY J., MARTIN J.C., DANIEL A.S., FLINT H.J.;				
RT	"pockerin-like sequences in cellulases and xylanases from the rumen				
RT	cellulolytic bacterium Ruminococcus flavefaciens."				
RL	FEMS MICROBIOL. LETT. 149:213-219(1997).				
RN	[2]				
RP	SEQUENCE OF 1-465 FROM N.A.				
RC	STRAIN=17;				
RA	CUNNINGHAM C., MCPHERSON C.A., MARTIN J.C., HARRIS W.J., FLINT H.J.;				
RL	SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.				
DR	EMBL; Z83304; Z285019; -.				
DR	PFAM; PF00150; cellulase; 1.				

	DB	13	VAFAGSVSAATLPTAK-EVQA--K-MGMFNTGNSME-VPNSTLNGNPPYTPQLDLSV	67
	QY	16	VSEFAVSSANEDVKTLDIQSIVRDMQPGWNLGNTDFAVGDQDETANGNPRTRELIERI	75
RESULT	11			
ID	O06842	PRELIMINARY;	PRT;	924 AA.
AC	O06842			
DT	01-JUL-1997	(TREMBUREL. 04, CREATED)		
DT	01-JUL-1997	(TREMBUREL. 04, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBUREL. 08, LAST ANNOTATION UPDATE)		
DE	B14,	PUTATIVE POLYGLUTAMINASE, B-1.4-ENDOGLUCANASE AND MANNANASE		
DE	GENES,	COMPLETE CDS.		
OS	PREVOTELLA RUMINICOLA (BACTEROIDES RUMINICOLA).			
OC	BACTERIA; CYTODHAGALES; BACTEROIDACEAE; PREVOTELLA.			
RN	[1]			
RC	STRAIN=B14;			
RC	RUSSELL J.B., DR GARDNER R.G., WELLS J.E., FIELDS M.W., WILSON D.B.;			
RC	SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.			
DR	EMBL; U96771; G2130574; -			
DR	PFAM; PF00150; cellulase; 1.			
SQ	SEQUENCE	924 AA; 103256 MW; CC043BE0 CRC32;		
		Query Match	10.6%; Score 447; DB 2; Length 924;	
		Best Local Similarity	29.2%; Pred. No. 2.53e-64;	
		Matches	79; Conservative	69; Mismatches 108; Indels 15; Gaps 11;
DB	611	PVATYETFWGPETTDMMTEFLMGNGFNARIPVTWYEHMDAEGN--VDEAWMRVKAIV	668	
QY	54	AVGODEFANGNPVRTRELIERIADGYKSIRIPVTWENRIGGAPDYDPQFLNRVDEVV	113	
		Sequence from N.A.		
DB	669	EYAMNAGLYAIRVVHHDTAAAGSGAWIKADTDVYAATKEFKLWTQIANALADYDQHLLF	728	
QY	114	QWALEEDLYVMINLHDS--W-LWIYEMEHNYGVMAKYRSLEQSLNHFDYPTKLME	169	
DB	729	EGYMEDLGNNSDPEPKASGEALNNAQFPVDRAVATGGGNATRNLIINTVYAAKGNN	788	
QY	170	ESVNEP-KFSQNWGEI-RENHALLDDLTTFEIVRSGGGNDIRPLVPTMETATSOP	227	
DB	789	VLNMELPDAVNN-HLVQVHSVPWNF-FNTKT-TW-DSCHTNLTIFALSALKFTT	844	
QY	228	LNNLYOTIDKLDPLNIATVTHYGFVPSVNIAGTYFEEDSKREIIETFRVHHFFVA	287	
DB	845	--IPYIIGXETHGESDISVSXSSPAEXIKL	873	
QY	288	RGIPVVLGEGLLG-FDKHTGVIQQGKLF	317	
RESULT	12			
ID	O47916	PRELIMINARY;	PRT;	519 AA.
AC	O47916			
DT	01-NOV-1996	(TREMBUREL. 01, CREATED)		
DT	01-NOV-1996	(TREMBUREL. 01, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBUREL. 08, LAST ANNOTATION UPDATE)		
DE	ENDOGLUCANASE CELG.			
OS	FIBROBACTER SUCCINOGENES (BACTEROIDES SUCCINOGENES).			
OC	BACTERIA; FIBROBACTER GROUP; FIBROBACTER.			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=S85;			
RC	MEDLINE; 97017599.			
RT	IYO A.H., FORSBURG C.W.;			
RT	"Endoglucanase G from Fibrobacter succinogenes S85 belongs to a class			
RT	of enzymes characterized by a basic C-terminal domain."			
RL	CAN. J. MICROBIOL. 42:934-943(1996).			
DR	EMBL; U33887; G1022698; -			
DR	PFAM; PF00150; cellulase; 1.			
SQ	SEQUENCE	519 AA; 56848 MW; FOC16BDC CRC32;		
		Query Match	9.4%; Score 395; DB 2; Length 519;	
		Best Local Similarity	28.2%; Pred. No. 9.06e-54;	
		Matches	70; Conservative	68; Mismatches 94; Indels 16; Gaps 14;

```

RN  SEQUENCE FROM N.A.
RP  STRAIN-M-777;
RC  MEDLINE; 97390413.
RA  IZU H., IZUMI Y., KURUME Y., SANO M., KONDO A., KATO I., ITO M.;
RT  "Molecular cloning, expression, and sequence analysis of the
RT  endoglycoceramidase II gene from Rhodococcus species strain M-777.";
RL  J. BIOL. CHEM. 272:19846-19850(1997).
DR  EMBL; U39554; G2337906; -.
KW  HYDROLASE; GLYCOSIDASE.
SQ  SEQUENCE 490 AA; 52750 MW; F4E87E0A CRC32;

Query Match      3.7%; Score 156; DB 2; Length 490;
Best Local Similarity 29.8%; Pred. No. 3.51e-08;
Matches 34; Conservative 31; Mismatches 40; Indels 9; Gaps 9;

Db  29 AASGSGSGTAL-TPSYLKD-DGGRSILRGFNFTASSAKSAPDGMPOFTADLAREYAD 86
QY  :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
21 PAVSSANEDVKTLDIQSVYRDMQPGWNLG-NTFDVAGQDETAW-CNPRVTR-ELIERIAD 77
Db  87 MGTNFRFLISWRS-VEPAPG-VYDQYLDYRVEDRVGWYAERG-YKVMMDMHQD 137
QY  :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
78 EGYKSIRIPVTWENRIGGAPDYPIDPQFLNRVDEVVQWALEEDLY-VMINLHHD 130

RESULT 15
ID  Q12539      PRELIMINARY;      PRT;      419 AA.
AC  Q12539;
DT  01-NOV-1996 (TREMBLREL. 01, CREATED)
DT  01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT  01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE  EXO-1,3-BETA-GLUCANASE PRECURSOR (EC 3.2.1.58)
DE  (GLUCAN 1,3-BETA-GLUCOSIDASE) (EXO-1,3-BETA-GLUCOSIDASE).
GN  ABEXG1.
OS  AGARICUS BISPORUS (COMMON MUSHROOM).
OC  EUKARYOTA; FUNGI; BASIDIOMYCOTA; HYMENOMYCETES; AGARICALES;
OC  AGARICACEAE; AGARICUS.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-HORST U1;
RX  MEDLINE; 96304295.
RA  VAN DE RHEE M.D., MENDES O., WERTEN M.W.T., HUIZING H.J.,
RA  MOOIBROEK H.;
RT  "Highly efficient homologous integration via tandem exo-beta-1,
RT  3-glucanase genes in the common mushroom, Agaricus bisporus.";
RL  CURR. GENET. 30:166-173(1996).
DR  EMBL; X92961; G1064880; -.
KW  SIGNAL; HYDROLASE; GLYCOSIDASE.
FT  SIGNAL 1 22 POTENTIAL.
SQ  SEQUENCE 419 AA; 46682 MW; 05344CCA CRC32;

Query Match      3.3%; Score 141; DB 3; Length 419;
Best Local Similarity 32.4%; Pred. No. 8.25e-06;
Matches 22; Conservative 19; Mismatches 23; Indels 4; Gaps 4;

Db  85 W-NTWITEEDFARIAAGLNHRLPIGYWAFVAAAGEPY-IQGG-LPFLERKAVTWAQNHN 141
QY  :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
62 WGNPRVTRIELIERIADEGYSKIRIPVT-WENRIGGAPDYPIDPQFLNRVDEVVQWALEED 120
Db  142 LKLIIDLH 149
QY  :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
121 LYVMINLH 128

```

Search completed: Fri Jun 4 10:03:56 1999
Job time : 126 secs.

MPARLH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm

on: Fri Jun 4 10:05:47 1999; MasPar time 22.00 Seconds
554.952 Million cell updates/sec

Tabular output not generated.

Title: >US-08-945-574-2
Description: (1-574) from US08945574.pep
Perfect Score: 4235
Sequence: 1 MKWKMVMYLAIVLYVSEFA.....GNQVTGIAQTNSKNKK 574

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 2126608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 36.906; Variance 180.757; scale 0.204

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	4235	100.0	574	19	W00383	0.00e+00
2	2557	60.4	566	3	R13228	9.20e-208
3	597	14.1	406	2	R08199	Neutral cellulase gen
4	535	12.6	452	9	R49102	Translated sequence o
5	531	12.5	800	9	R47496	Translated sequence o
6	517	12.2	471	32	W56742	Orpinomyces cellulase
7	335	7.9	360	26	W34566	Thermotoga OCI/4V end
8	326	7.7	360	36	W49870	Thermotoga OCI/4V end
9	155	3.7	461	23	W10210	Mature endoglucoceram
10	156	3.7	490	23	W10209	Full length endoglyco
11	150	3.5	418	32	W57421	Amino acid sequence o
12	141	3.3	418	14	R79540	Endoglucanase-II prot
13	141	3.3	418	14	R77264	T. longibrachiatum en
14	141	3.3	418	18	W02032	Trichoderma endogluc
15	119	2.8	429	17	R88407	Beta-(1,6)-endoglucan
16	113	2.7	454	26	W34559	Thermococcus AEDII12R

17	113	2.7	454	36	W49863	Thermococcus AEDII12R
18	106	2.5	269	31	W28052	Amino acid sequence o
19	104	2.5	400	25	W22521	Bacillus agaradherens
20	104	2.5	400	25	W23601	Bacillus agaradherens
21	104	2.5	400	32	W57431	Bacillus agaradherens
22	104	2.5	462	32	W57433	Cloned alkaline endog
23	107	2.5	510	8	R47456	Truncated xylanase (X
24	107	2.5	551	8	R47455	Truncated xylanase (X
25	107	2.5	607	8	R47454	Xylanase (XYLA).
26	107	2.5	644	9	R44529	PNPX30 xylanase.
27	106	2.5	645	34	W29653	Human secreted protei
28	106	2.5	1969	36	W72419	Rice bacterial leaf b
29	100	2.4	176	19	R38153	Thermostable inorgani
30	100	2.4	272	36	W69432	Human secreted protei
31	103	2.4	551	24	W18790	Corrected Bacillus la
32	98	2.3	249	37	W72670	Canine herpes virus p
33	98	2.3	249	26	W23008	Canine herpesvirus un
34	96	2.3	420	13	R75754	BAV3 E1B 56K protein.
35	97	2.3	421	26	W34556	Staphylothermus marin
36	97	2.3	421	36	W49860	Staphylothermus marin
37	96	2.3	584	3	R14670	Truncated poly Ig-rec
38	96	2.3	773	20	W03177	Rabbit poly-immunoglo
39	96	2.3	869	10	R53732	S. cerevisiae Plci pr
40	96	2.3	883	14	R71415	E. coli PEPC (wild-ty
41	96	2.3	936	9	R52579	Recombinant collagena
42	98	2.3	1008	33	W63721	C. histolyticum CHCI
43	97	2.3	2721	2	P70647	Sequence of N-termina
44	94	2.2	411	21	W12379	P300-CelB fusion cons
45	94	2.2	412	21	W12380	P300-CelB fusion cons

ALIGNMENTS

RESULT 1
ID W00383 standard; Protein; 574 AA.
AC W00383;
DE 31-JAN-1997 (first entry)
DE Bacillus cellulase BCE 113.
KW Cellulase; BCE 113; detergent; surfactant; laundry;
KW tensile strength; antipilling;
OS Bacillus sp. strain CBS 669.93.
FH Key Location/Qualifiers
FT peptide 1..26
FT protein /label= Sig_peptide
FT protein /label= Mat_protein
FT /note= "the 63 kDa mature protein is used in
detergent compns."

W09634092-A2
31-JAN-1997

26-APR-1996; E01755.
28-APR-1995; EP-201115.
12-MAR-1996; US-614115.
PA (GENV) GENENCOR INT INC.
PI Kottwitz B, Lenting HBM, Maurer K, Van Beckhoven RFWC;
PI Van Solingen P, Weiss A;
DR WPI; 96-497624/49.
DR N-PSDB; T41849.

Cellulase with low ratio of tensile strength loss to antipilling
properties - used in detergent composition which provides
anti-greying, softening, anti-wrinkling and colour protection to
fabrics
Claim 10; Fig 4; 33pp; English.
A novel cellulase (W00383), designated BCE 113, of alkali-tolerant
Bacillus sp. strain CBS 669.93 has a tensile strength loss to
antipilling ratio below 1. It can be isolated from CBS 669.93
fermentation broth or expressed at high levels in transformed host
cells utilising an isolated gene sequence (T41849). Cellulase BCE
113, and similarly isolated cellulase BCE 103 (see also W00382),
show good activity at alkaline pH, and can be used in laundry
detergent compns. to provide anti-greying, softening, anti-
wrinkling and colour protection to fabrics.
SQ Sequence 574 AA;

Query Match 100.0%; Score 4235; DB 19; Length 574;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00; Mismatches 0; Indels 0; Gaps 0;
 Matches 574; Conservative 0;

Db 1 mkwmksmwlaavlvsfapavssanedvklldiqsvvrdmqgwnlgnftdavgqdet 60
 |||||
 QY 1 MKWKSVMWLAVALVVSFAPAVSSANEDVKTLDIQSVVRDMPQGNLGNFTDVGQDET 60
 |||||

Db 61 awgnprvtrelrieriadeqysiripvtwenriggapydipqflnrdevvqwaaleed 120
 |||||
 QY 61 AWGNPRVTRELIERIADEQYSIRIPVTWENRIGGAPYDIPQFLNRDEVVQWALEED 120
 |||||

Db 121 lyminlhhsdswliwemehnyngmakysrlweqlshhfkdyptkklmfesvnepkfsqn 180
 |||||
 QY 121 LYMINLHHSWSLWIWEMEHNYNGMAKYSRLWEQLSHHFKDYPTKLMFESVNEPKFSQN 180
 |||||

Db 181 wgeirenhalldlntvffeiavqsqgqndirplvptmetatsqpllnnlyqtidkld 240
 |||||
 QY 181 WGEIRENHALLDLNLTNVTFFEIVRQSGQNDIRPLVPTMETATSQPLNNLYQTIDKLD 240
 |||||

Db 241 dplnatvhygfwfsvniagytfrfeedskreietfdrvhhtfvargipvlgfcll 300
 |||||
 QY 241 DPNLIATVHYGFWFVSNIAGYTRFEEDSKREIETFDRVHHTFVARGIPVLGFECLL 300
 |||||

Db 301 gfdkhtvgioqgeklkffeylihlhlnrdithmlwdngqghfnrhtyewydeelfdmrlas 360
 |||||
 QY 301 GFDKHTVGIOQGEKLKFFEYLIHLHNRDITHMLWDNGQGHFNRTYEWYDEELFDMRLAS 360
 |||||

Db 361 wgrssvaesnfylkggdriadatvqlqhgnetlglqangqrtpqdyelangerltv 420
 |||||
 QY 361 WGRSSVAESNFYILKGGDRIADATVTLQHGNETLGLQANGQRTPQDYELANGERLTV 420
 |||||

Db 421 kahvlsaiagsgtlgtngmvtcaefnrgadwhfrvntvrtplqstqghvsnfsipasnfg 480
 |||||
 QY 421 KAHVLSAIAAGSGTLGTNGMVTAEFNRGADWHFRVNTVTPVLTQSTQGHVSNFSIPASNFG 480
 |||||

Db 481 nslatmeavydngnagpddwtskfeqyafspysydttheiklteaffrevrdgevrftfh 540
 |||||
 QY 481 NSLATMEAVYDNGNAGPDDWTSKFEQYAFSPSYDTTHEIKLTEAFFREVDRDGEVRFTFH 540
 |||||

Db 541 fwsgeivnytiikngnqvgtgiaaattasknknk 574
 |||||
 QY 541 FWSGEIVNYTIKNGNQVTGIAAATTNSKNKNK 574
 |||||

RESULT 2
 ID R13228 standard; Protein; 566 AA.
 AC R13228;
 DT 14-OCT-1991 (first entry)
 DE Endoglucanase encoded by endo2 gene.
 OS Cellulase activity; detergent.
 OS Bacillus spp. NCIMB 40250.
 FH Key Location/Qualifiers
 FT cleavage_site 30..31
 FT peptide 1..30
 FT /label= signal peptide
 FT peptide 31..566
 FT /label= mature peptide
 FT W09110732-A.
 PN 25-JUL-1991.
 PD 18-JAN-1991; DK0013.
 PE 19-JAN-1990; DR-000164.
 PR (NOVO) NOVO NORDISK A/S.
 PA Jorgensen PL, Schulein M, Hansen C;
 PI N-PSDB; Q13002.
 DR Enzyme exhibiting cellulase activity from Bacillus sp. - is an
 PT endo-glucanase, esp. useful for harshness redn. of cotton-contg.
 PT fabrics.
 PS Claim 1; Page 80; 96pp; English.
 CC The enzyme is produced by transforming plasmid pPL382 to B-subtilis
 CC DnJ85 spp. and exhibits an endoglucanase activity of at least 10

CC (pref. at least 25) carboxymethyl cellulose (CMC) endoase units per
 CC mg total protein under alkaline conditions. It is especially useful
 CC as a cellulytic agent and has been found to be more stable during
 CC washing (60 mins. at 40 deg.) in the presence of conventional
 CC detergents than a commercial cellulase preparation. It may also
 CC show increased storage stability in liq. detergents contg.
 CC proteases. The sequence was deduced from the DNA (Q13002), it is a
 CC product of the endo2 gene. See also R13227 and R13229.
 SQ Sequence 566 AA;

Query Match 60.4%; Score 2557; DB 3; Length 566;
 Best Local Similarity 57.6%; Pred. No. 9.20e-208;
 Matches 326; Conservative 117; Mismatches 119; Indels 4; Gaps 4;

Db 2 kkrsskvlslavallaaevpnaalaaappsamqsyveamqpgwnlgnslavgade 61
 |||||
 QY 1 MKWKSVMWLAVALVVSFAPAVSSAN-EDVKTLDIQSVVRDMPQGNLGNFTDVGQDE 59
 |||||

Db 62 tlargnpritrkellqnaaqysiripvtwdshigaapnygieaaylnrvqevvqwal 121
 |||||
 QY 60 T-ANGNPRVTRRELIERIADEQYSIRIPVTWENRIGGAPYDIPQFLNRDEVVQWALE 118
 |||||

Db 122 anlyminhhdswliwiskmesqhdqlarynaiwtgiantkfnspsklmfesvneprft 181
 |||||
 QY 119 EDLIVMINLHSDSLWIWYEMEHNYNGMAKYSRLWEQLSHHFKDYPTKLMFESVNEPKFS 178
 |||||

Db 182 dg-gttdeakqkmldeinvssfnivrsngqgnatrpvlstleasptqermtalynmt 240
 |||||
 QY 179 QNWGEIRE-NHALLDLNLTNVTFFEIVRQSGQNDIRPLVPTMETATSQPLNNLYQTID 237
 |||||

Db 241 klnkniatvhygfwfsvniagytfrfeedskreietfdrvhhtfvargipvlgfcll 300
 |||||
 QY 238 KLDOPNLATVHYGFWFVSNIAGYTRFEEDSKREIETFDRVHHTFVARGIPVLGFE 297
 |||||

Db 301 gllgfdkntvgiegeklkffeffaqyvkksissmlwdngqghfnrhtyewydeelfdm 360
 |||||
 QY 298 GLLGFDKHTVGIOQGEKLKFFEYLIHLHNRDITHMLWDNGQGHFNRTYEWYDEELFDM 357
 |||||

Db 361 kasvtgrsstassdlihwkgtavkdtstvqlnlnpntslsvngttlksgtdylnssr 420
 |||||
 QY 358 RASWGRSSVAESNFYILKGGDRIADATVTLQHGNETLGLQANGQRTPQDYELNGER 417
 |||||

Db 421 ltfkasqltkits:lgkigvnativtkkfnrgadwhfrvntvrtplqstqghvsnfsip 480
 |||||
 QY 418 LTVKAHVLSAIAAGSGTLGTNGMVTAEFNRGADWHFRVNTVTPVLTQSTQGHVSNFSIPAS 477
 |||||

Db 481 lngdqlatmeavyngnagphnwtskfettfsspayregkiklqqaaffnevdttvtl 540
 |||||
 QY 478 FNGSLATMEAVYDNGNAGPDDWTSKFEQYAFSPSYDTTHEIKLTEAFFREVDRDGEVR 537
 |||||

Db 541 kfgfwsggeivnytiiksgstvtgtas 566
 |||||
 QY 538 TFHFWSGEIVNYTIKNGNQVTGIAA 563
 |||||

RESULT 3
 ID R08199 standard; protein; 406 AA.
 AC R08199;
 DT 01-MAR-1991 (first entry)
 DE Neutral cellulase gene product.
 DE Cellulose.
 KW Ruminococcus albus.
 OS Key Location/Qualifiers
 FT misc_difference 44..58
 FT /label= N-terminal deletion
 FT 44..406
 FT protein
 FT J02265486-A.
 PN 30-OCT-1990.
 PD 07-APR-1989; 086714.
 PE 07-APR-1989; JP-086714.
 PR (SHIM/) SHIMIZU S.
 PA Shimizu S;
 PI WPI; 90-366319/49.

```

Db 108 ngafvlnlhthetwnhafs-e-tldtakellekiwsgiaefkdydehllifeglnepkrn 165
      : :::::||||: | : : | ::: |||| |:::||||
Qy 119 EDLYVMINLHSDSLWIYEMEHNNYGVMAKYSRLWEQLSNHFKDYPPTKLMFESVNEP-KF 177
      : :::|:::||||: | : : | ::: |||| |:::||||
Db 166 dtpvewtgqdegwdav-namnavflktirssgnnpkrrhlmpypaaacnensfnk-f- 222
      : :::|:::||||: | : : | ::: |||| |:::||||
Qy 178 -SQ-NW-GEIRENHALLDNLNTVFFFIIVRQSGQNDIRPLVLTMETATSQPLLNNLYQ 234
      : :::|:::||||: | : : | ::: |||| |:::||||
Db 223 ifpe-dddkviasvhaypynfalnngegavdkfdaagkdkdlewnlnlmkkrfvdgipm 281
      : :::|:::||||: | : : | ::: |||| |:::||||
Qy 235 TIDKLDPNLIATVHYGVFPFSSVNI-AGYT-REEDSKREIETETFRVHHTFVARGIPV 292
      : :::|:::||||: | : : | ::: |||| |:::||||
Db 282 ilgygam--nrnd--ee-eratwafymekvtamgvpqvwndng 321
      : :::|:::||||: | : : | ::: |||| |:::||||
Qy 293 VLGEFGLGDPKHTGVIQGEKLFEEYLIHHLNERDITHMLWDNG 338
      : :::|:::||||: | : : | ::: |||| |:::||||

RESULT 5
ID R47496 standard; Protein; 800 AA.
AC R47496;
DC 21-JUL-1994 (first entry)
DE Translated sequence of domains I and II of celd cDNA in clone pCNP4.
KW Cellulase; celd; pCNP4; anaerobic rumen.
OS Neocallimastix patriciarum.
FH Key
FT region 1..37
FT /label= N-terminus of beta-galactosidase alpha-
FT peptide
FT misc_difference 38..42
FT /note= "derived from 5' oligo linker"
FT W09400578-A.
PD 06-JAN-1994.
PF 24-JUN-1993; AU0307.
PR 24-JUN-1992; AU-003096.
PA (GSIR ) COMMONWEALTH SCI & IND RES ORG.
PI Aylward JH, Gobius KS, Orpin CG, Xue GP;
DR WPI; 94-026214/03.
DR N-PSDB: Q55036.
PT Cloning of cellulase clones from anaerobic rumen - by isolating
PT mRNA from culture, converting to cDNA etc. fungi, producing
PT enzymes useful in food processing etc., and DNA for modifying
PT rumen or silage bacteria
PS Claim 10; Page 41-45; 71pp; English.
CC Clone pCNP4.1 encodes celd. It was derived as follows. N.
CC patriciarum was used to construct a cDNA library in ZAPII (in E.
CC coli). Transformants were selected for enzyme activity using
CC selective media. 11 colonies were positive, and of these 10 had the
CC same restriction pattern, and the longest of these was designated
CC celd (pCNP4.1)(Q55036)(R47496,R49102). A similar process was used
CC to isolate the xylanase clone pNX-tac (Q55037,R47497). An enzyme
CC composition contg. celd and xylanase is claimed.
SQ Sequence 800 AA;

Query Match 12.5%; Score 531; DB 9; Length 800;
Best Local Similarity 32.9%; Pred. No. 1.49e-32;
Matches 83; Conservative 72; Mismatches 85; Indels 12; Gaps

Db 84 etcwgnpkttedmfkvlmdnqfnvfrlptbtwsglhfgeapdykineklwkrvheivdyk 143
      : :::|:::||||: | : : | ::: |||| |:::||||
Qy 59 ETANGNPRVTLERITAEQYKYSIRIPVTWENRIGGAPDYPIDPQFLNRVDEVVQWALE 118
      : :::|:::||||: | : : | ::: |||| |:::||||
Db 144 ngafvlnlhthetwnhafs-e-tldtakellekiwsgiaefkdydehllifeglnepkrn 201
      : :::|:::||||: | : : | ::: |||| |:::||||
Qy 119 EDLYVMINLHSDSLWIYEMEHNNYGVMAKYSRLWEQLSNHFKDYPPTKLMFESVNEP-KF 177
      : :::|:::||||: | : : | ::: |||| |:::||||
Db 202 dtpvewtgqdegwdav-namnavflktirssgnnpkrrhlmpypaaacnensfnk-f- 258
      : :::|:::||||: | : : | ::: |||| |:::||||
Qy 178 -SQ-NW-GEIRENHALLDNLNTVFFFIIVRQSGQNDIRPLVLTMETATSQPLLNNLYQ 234
      : :::|:::||||: | : : | ::: |||| |:::||||
Db 259 ifpe-dddkviasvhaypynfalnngegavdkfdaagkdkdlewnlnlmkkrfvdgipm 317
      : :::|:::||||: | : : | ::: |||| |:::||||
Qy 235 TIDKLDPNLIATVHYGVFPFSSVNI-A-GYT-REEDSKREIETETFRVHHTFVARGIPV 292
      : :::|:::||||: | : : | ::: |||| |:::||||

```

Db 318 ilgeygamrldn 329
:||||:|:|:
QY 293 VLGEFGLGFDDK 304

RESULT

ID W56742 standard; Protein; 471 AA.

AC W56742; 6

DT 14-SEP-1998 (first entry)

DE Orpinomyces cellulase celB.

KW Cellulase; endoglucanase; cellobiohydrolase; celB.

OS Orpinomyces sp. strain PC-2.

PN W09814597-A1.

PD 09-APR-1998.

PF 03-OCT-1997; U18008.

PR 04-OCT-1996; US-027883.

PA (UYGE-) UNIV GEORGIA RES FOUND INC.

PI Chen H, Li X, Ljungdahl LG;

DR WPI; 98-240096/21.

DR N-PSDB; V29477.

PT New recombinant DNA encoding Orpinomyces cellulase protein - useful for, e.g. producing recombinant Orpinomyces cellulase in host cell

PS Claim 1; Page 38-40; 69pp; English.

CC This polypeptide comprises cellulase celB of Orpinomyces sp. strain CC-2, an anaerobic bovine rumen fungus Orpinomyces sp. PC-2. Its amino acid sequence was deduced from an isolated cDNA clone (see W29477). CelB has cellobiohydrolase activity, with highest activity at pH 5.2-6.2 and 50 degC. CelA (see W56738) and celC (see W56739) cellulolytic enzymes of Orpinomyces sp. PC-2 are also provided. Recombinant DNA molecules encoding Orpinomyces cellulase proteins are claimed, as well as recombinant cells selected from

CC Saccharomyces cerevisiae, Escherichia coli, Aspergillus, CC Trichoderma reesei, Pichia, Penicillium, Streptomyces or Bacillus, CC and a method for producing recombinant cellulase by culturing these CC host cells.

CC Sequence 471 AA;

SQ

Query Match

Best Local Similarity 12.2%; Score 517; DB 32; Length 471;

Matches 104; Conservative 99; Mismatches 123; Indels 24; Gaps 19;

Db 1 mkfinslglviagcaamrniss-kelvkeltigwslgntldascvetlnyskdtas 59

QY 1 MKWKMSVWLAVLVGVFAPAVSANEDVKTLDIQ-SYVRDMQPGWNLGNTFD-AVGQD 58

Db 60 etcgvnvtktgelyklsdignfrpttwtgfhgdpdykisdvwmkrvhevvdvaln 119

QY 59 ETANGNPRVTRELLIERADEGYKIRIPVTWENRIGGAPDPIDPQFLNRVDEVVQWALE 118

Db 120 tggvialinhethw--nyafgknllesakkilvakiwqtaafgdydehlifegmneprkv 177

QY 119 EDLYVMINLHSDSLWIYEMEHNYNGVMKYRSLWEQLSNHFKDYPYTKLMFESVNEP-KF 177

Db 178 gdpawtgqdgqevn-fvnemnalfvktirgtgnnanrhlmptyaasvndgsinn-fk 235

QY 178 SQ--NW-GEIRENHALLDLNTVFEIVROSGQNDIRPLVLTMETATSQPLLNLIYQ 234

Db 236 ypng-dd-kvlsvlshyspynfalnpggalsnfy-dgn-eidwvmtinsfskigpiv 291

QY 235 TIDKLDPNLIATVHYGVFWPFVSNIA-G-YTRFEEDSKREIETFDVHVHTFVARGIPV 292

Db 292 lgeff--vamrdrn--ed-drerwqeyikkatalgipcvlwdng-yfe 334

QY 293 VLGEFGLGFDDKHTGVYQGEKLAFFFEYLIHLNERNDRITHMLWDNGQHFN 342

RESULT

ID W34566 standard; Protein; 360 AA.

AC W34566;

DT 12-MAR-1998 (first entry)

DE Thermotoga OCl/4V endoglucanase.

KW Glycosidase; thermostable; textile; food processing; pharmaceutical;

KW detergent; baking; industry; Thermococcus; Staphylothermus;
OS Pyrococcus; glucose; soluble oligosaccharide; endoglucanase.
PN Thermotoga sp.
PD W09723417-A1.
PF 17-JUL-1997.
PR 10-JAN-1997; U00092.
PR 13-SEP-1996; US-712612.
PR 11-JAN-1996; US-583787.
PA (RECO-) RECOMBINANT BIOTRANSFORMATION INC.
PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
DR WPI; 97-372858/34.
DR N-PSDB; T93690.
PT New thermostable glycosidase(s) - from Thermococcus, Staphylothermus
PT and Pyrococcus, used in the textile, food processing,
PT pharmaceutical, detergent and baking industries
PS Claim 4; Fig 13; 82pp; English.
CC The present sequence represents endoglucanase isolated from Thermotoga.
CC The enzyme or its encoding nucleic acid sequence is used for generating
CC glucose from soluble oligosaccharides. The enzyme can be used in the
CC food processing, pharmaceutical, textile, detergent and baking
CC industries. The enzyme is also used to treat lactose intolerance, as a
CC diagnostic reporter molecule, in corn wet milling or in the fruit juice
CC industry. The enzymes can be used to hydrolyse guar gum to remove
CC non-reducing terminal mannose residues. The nucleic acids encoding the
CC enzyme may be used to generate probes to identify similar sequences.
SQ Sequence 360 AA;

Query Match 7.9%; Score 335; DB 26; Length 360;

Best Local Similarity 42.3%; Pred. No. 1.82e-16;

Matches 55; Conservative 28; Mismatches 37; Indels 10; Gaps 6;

Db 62 nignalea-pfegawgv-rieheylfeinkrgfdsvripirwsahisekppydidrnl 118

QY 47 NLGNTDVGQDETAGNPRVTRELLIERADEGYKIRIPVTWENRIGGAPDPIDPQFL 106

Db 119 ervnhvdralenltviinthfee-l--ygepdkygdvive---ivrqakfkdkype 172

QY 107 NRVDVGVQWALEEDLYVMINLHH-DSWLWYIEMEHNYNGVMKYRSLWEQLSNHFKDPT 165

Db 173 nlffeinyneep 182

QY 166 KLMFESVNEP 175

RESULT

ID W49870 standard; Protein; 360 AA.

AC W49870;

DT 21-DEC-1998 (first entry)

DE Thermotoga OCl/4V endoglucanase.

KW Glycosidase; OCl/4V; thermostable enzyme; oligosaccharide; glucose;

OS sugar; baking; textile; detergent; endoglucanase.

PN Thermotoga sp. strain OCl/4V-33Gp1.

PD 11-JUN-1998.

PF 08-DEC-1997; U22623.

PR 10-OCT-1997; US-949026.

PR 06-DEC-1996; US-056916.

PA (DIVE-) DIVERSA CORP.

PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;

DR WPI; 98-362407/31.

DR N-PSDB; V36919.

PT Glycosidase enzymes from organisms of the genera Staphylothermus,
PT Pyrococcus and Thermococcus - for deriving sugar from
PT oligosaccharides, useful in the e.g. food processing, textile or
PT baking industries

PS Claim 1; Fig 13a-b; 92pp; English.

CC This is the amino acid sequence of endoglucanase OCl/4V, deduced

CC from a polynucleotide (see V36919) of a Thermotoga sp. OCl/4V clone
CC (33Gp1) that grows optimally at 75 degC. The sequence shows 65%
CC amino acid identity to an endo-1,4-beta-endoglucanase of
CC Caldocellum thermocellum. The invention provides 18 polynucleotides
CC (see V36907-24) coding for thermostable glycosidases (see W49858-75)
CC having glucosidase, alpha-galactosidase, beta-galactosidase,

KW	Endoglycoceramidase; glycolipid; sugar chain engineering.
OS	Rhodococcus sp. M-777.
FH	Key Location/Qualifiers
FT	peptide 1..29
FT	/label= Signal_peptide
FT	protein 30..490
FT	/label= Mature_protein
EP	-751222-A2.
PN	02-JAN-1997.
PD	28-JUN-1996; 110523.
PF	29-JUN-1995; JP-188465.
PR	(TAKI) TAKARA SHUZO CO LTD.
PA	Ito M, Izu H, Izumi Y, Kato I, Kurome Y, Sano M;
PI	WPI; 97-054679/06.
DR	N-PSDB; T50943.
DR	New DNA encoding endo-glyco-ceramidase - useful in structural and
PT	functional analysis of glyco-lipid(s)
PS	Claim 1; Page 11-13; 25pp; English.
CC	The sequences given in W0209-10 represent full length and mature
CC	endoglycoceramidase respectively, which is useful in structural
CC	and functional analysis of glycolipids in sugar chain engineering.
CC	Recombinant endoglycoceramidase can be produced with high purity at
CC	relatively low cost using the DNA sequences encoding these proteins.
CC	Sequence 490 AA;
SQ	
Query Match	3.7%; Score 156; DB 23; Length 490;
Best Local Similarity	29.8%; Pred. No. 1.04e-02;
Matches	34; Conservative 31; Mismatches 40; Indels 9; Gaps 9;
Db	29 aaagsgsgsgtal-tpsyikd-ddgrslilrfgntassaksapdgmpqftadlarevad 86 : ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: : :::
QY	21 PAVSSANEDVKTLDIQSYYRDMQPGNLNG-NTFDVAGQDETAW-GNPRVTR-ELIERIAD 77
Db	87 mgtnfurflwrs-vepapg-vyddyqlrdvredrvgywaerg-ykvmldmhqd 137 :
QY	78 EGYKSIRIPWTENRIGGAPDYPIDPQLNRVDVVQWALEEDLY-VMLNLHD 130
RESULT	11
ID	W57421 standard; Protein; 418 AA.
AC	W57421;
DT	02-SEP-1998 (first entry)
DE	Amino acid sequence of the specification.
KW	Regulatory sequence; cellulase cbhl gene; mass production;
KW	Humicola insolens; endo-glucanase NCE4.
OS	Trichoderma viride.
FH	Key Location/Qualifiers
FT	Peptide 1..21
FT	/note= "signal peptide"
FT	Protein 22..418
FT	/note= "mature protein"
FN	WO9811239-A1.
PD	19-MAR-1998.
PF	16-SEP-1997; J03268.
PR	13-SEP-1996; JP-243695.
PA	(MELJ) MELJI SEIKA KAISHA LTD.
PI	Aoyagi K, Moriya T, Murakami T, Sumida N, Watanabe M;
DR	WPI; 98-250959/22.
DR	N-PSDB; V2B597.
PT	Regulatory sequence for Trichoderma viride derived cellulase cbhl
PT	gene - for producing Humicola insolens derived endo-glucanase
PS	Disclosure; Pages 48-51; 92pp; Japanese
CC	The present sequence appears in the specification. The specification
CC	describes a new regulatory sequence for Trichoderma viride derived
CC	cellulase cbhl gene and the establishment of a system for mass producing
CC	cellulase in moulds such as T. viride. As the regulatory sequence of
CC	cbhl genes originating in T. viride can highly express objective
CC	proteins, proteins such as cellulase can be expressed. An expression
CC	vector containing the regulatory sequence and Humicola insolens derived
CC	endo-glucanase NCE4 DNA was produced, and used to produce endo-glucanase
CC	at 15 grams per litre.
CC	Sequence 418 AA;
SQ	

DE	T. longibrachiatum endoglucanase EGII.
KW	Cellulase; cellulose; signal; catalytic core; cellulase binding;
KL	linker.
OS	Trichoderma longibrachiatum.
FH	Key Location/Qualifiers
FT	peptide 1..21
FT	/label= signal
FT	/note= "seq id no 32"
FT	domain 22..57
FT	/label= cellulose binding domain
FT	/note= "seq id no 7"
FT	region 58..91
FT	/label= linker
FT	/note= "seq id no 24"
FT	domain 92..418
FT	/label= catalytic core domain
FT	/note= "seq id no 16"
PN	W09516782-A.
PD	22-JUN-1995.
PF	19-DEC-1994; U14163.
PR	17-DEC-1993; US-169948.
PA	(GENV) GENECOR INT INC.
PI	Clarkson KA, Collier KD, Fowler T, Larenas E, Ward M;
DR	WFI; 95-231574/30.
DR	N-PSDB; Q91284.
PT	Pure, truncated fungal cellulase protein from Trichoderma - useful to
PT	reduce or eliminate dye, colourant or pigment back-staining or
PT	redemption in stone-washing or bio-polishing
PS	Claims 17, 41, 63, 64, 65; Figure 4; 105pp; English.
CC	Figure 4 depicts the genomic DNA and AA sequence of EGII derived
CC	from T. longibrachiatum. A truncated fungal cellulase which
CC	consists essentially of the AA sequence in SEQ ID no. 8 is claimed.
CC	Also claimed are DNA gene fragments encoding SEQ ID Nos: 12, 20
CC	and 16; 16 and 20; and 16, 20 and 12.
CC	Genes for EGI and EGII have been isolated from T. longibrachiatum
CC	and the protein domain structure has been confirmed (Penttilä, M.
CC	et al., 1986, Gene, 45, 253-263; Van Arsdeell, J.N. et al., 1987,
CC	Bio/Technology, 5, 60-64; and Salohelmo, M. et al., 1988, Gene
CC	63, 11-21).
SQ	Sequence 418 AA;
Query Match 3.3%; Score 141; DB 14; Length 418;	
Best Local Similarity 22.8%; Pred.No. 1.18e-01;	
Matches 23; Conservative 30; Mismatches 46; Indels 2; Gaps	
Db	142 vnedgmtifrlprgvylnnlgldstskysqdlvqgclsgaycivdiuhayar-w 200
QY	: : :: :: :: : : : : :: :: : : :: : :
QY	75 IADEGYSIRIPYTWNRRGGAPDPIDPQFLNRVDEVQWALEDLYMINLHHDSWLW 134
Db	201 nggilggpggttaftslsqaskyas-qsrwvfginnep 240
QY	: : : : :: : : : ::
QY	135 IYEMEHNYGVMAKYSLEWQLSNHFKDYPYTKLMFESVNEP 175
RESULT 14	
ID	W02032 standard; Protein; 418 AA.
AC	W02032:
DT	28-OCT-1996 (first entry)
DE	Trichoderma endoglucanase II.
KW	Endoglucanase II; EGII; cellulase; cellulose; denim;
KX	stonewashing; dye redeposition; backstaining.
OS	Trichoderma longibrachiatum.
FH	Key Location/Qualifiers
FT	peptide 1..21
FT	/label= Sig_peptide
FT	protein 22..418
FT	/label= Mat_protein
FT	domain 22..57
FT	/label= Cellulose_binding_domain
FT	region 58..91
FT	/label= Linker_region
FT	domain 92..418
FT	/label= Catalytic_core_domain

FT /note= "catalytic core domain is the preferred
FT domain for use in constructs of the
FT invention"

PN WO9623928-A1.
PD 08-AUG-1996.
PF 29-JAN-1996; U00977.
PR 01-FEB-1995; US-382452.
PA (GEMV) GENENCOR INT INC.
PI Clarkson KA, Collier KB, Fowler T, Larenas E, Ward M;
DR WPI: 96-371466/37.
DR N-PSDB: T32223.
PT Treatment of cellulose-contg. fabrics such as denim, e.g.
PT stone-washing - using truncated cellulase enzyme to increase
PT abrasion and give reduced redeposition of dye
PS Disclosure: Fig 4A-4C; 124pp; English.
CC The amino acid sequences for Trichoderma longibrachiatum
CC cellobiohydrolase I (CBHI) (W02022), CBHII (W02025), endoglucanase I
CC (EGI) (W02029), EGII (W02032), and EGIII (W02034) were deduced from
CC the respective genomic DNA sequences (T32220-24). The CBHI, CBHII,
CC EGI and EGII enzymes have catalytic core domains useful for reducing
CC dye redeposition (backstaining) on cellulose-contg. fabrics such as
CC denim, whilst maintaining or increasing abrasion during stonewashing.
CC Truncated enzymes comprising these catalytic core domains can be obtd.
CC by proteolysis of the complete enzyme or by inserting the appropriate
CC DNA fragment into a vector, using this to transform a Trichoderma
CC sp. host cell, and recovering the recombinant core domain.
SQ Sequence 418 AA;

Query Match 3.3%; Score 141; DB 18; Length 418;
Best Local Similarity 22.8%; Pred. No. 1.18e-01;
Matches 23; Conservative 30; Mismatches 46; Indels 2; Gaps 2;
Db 142 vnedgmtfirlpvvgvlynnlgnldstskyskylvggclslgacyclvdihnyar-w 200
QY 75 IADGKYSIRIPVTWENRIGGAPDPIDPQLNRVDEVVQWALEEDLYVMINLHDSWLW 134
Db 201 nggllgqgqptnaqftslwslaskyas-qsrwvfgimnep 240
QY 135 IYEMEHNYNGVMKYRSLWEQLSNHFKDYPTKLMFESVNEP 175

RESULT 15
ID R88407 standard; Protein: 429 AA.
AC R88407;
DT 05-JUL-1996 (first entry)
DE Beta-(1,6)-endoglucanase.
KW Trichoderma harzianum beta-(1,6)-endoglucanase; Aspergillus oryzae;
KW beta-glucan degradation; pustulanase.
OS Trichoderma harzianum.
PN WO9531534-A1.
PD 23-NOV-1995.
PE 11-MAY-1995; DK0189.
PR 11-MAY-1994; DK-000547.
PA (NOVO) NOVO-NORDISK AS.
PI Andersen LN, Christgau S, Dalboge H, Kauppinen MS;
PI Kofod LV, Olsen HS;
DR WPI: 96-010921/01.
DR N-PSDB: T09881.
PT DNA encoding beta-1-6-endoglucanase from Trichoderma harzianum -
PT useful, e.g., in prepn. of yeast extracts, as antifungal agent, in
PT cleaning compns., etc.
PS Claim 1; Page 35; 49pp; English.
CC Trichoderma harzianum (CBS 243.71) beta-(1,6)-endoglucanase may be
CC produced recombinantly, and has a molecular weight of about 50,000,
CC an apparent isoelectric point of 5.6, an optimal temperature of
CC 30-40 deg, an optimal pH of 4-6, preferably 5, and a specific
CC activity of about 100 U/mg. The enzyme, or compositions enriched
CC in it, are used to modify or degrade beta-glucans, particularly for
CC rupturing or lysing cell walls of microorganisms thereby enabling
CC recovery of desirable products produced by the microorganisms.
CC Typical applications include the production of protoplasts and
CC yeast extracts, use in wine and press juice making, as fungicides,
CC to remove excess dye from textiles, to remove moulds on coatings or

CC biofilms from surfaces, for cleaning dentures and removing plaque,
CC and in the extraction of mannoproteins from microbial cell walls.
SQ Sequence 429 AA;
Query Match 2.8%; Score 119; DB 17; Length 429;
Best Local Similarity 32.8%; Pred. No. 3.64e+00;
Matches 22; Conservative 17; Mismatches 24; Indels 4; Gaps 4;
Db 111 winpatvqsvhdvlgltiripigwysnaivdtasepfadgnlqlpyldavvqkaedlgi 170
QY 66 RVTRELIERIADEGKYSIRIPVT-WE-NRIGGAPDPVI-DPOF-LNRVDEVVQWALEEDL 121
Db 171 yviidlh 177
QY 122 YVMINLH 128
Search completed: Fri Jun 4 10:07:48 1999
Job time : 121 secs.

(™)

Result No.	Query			DB	ID	Description	Pred. No.
	Score	Match	Length				
1	2570	60.7	566	2	JH0218	cellulase (EC 3.2.1.4	0.00e+00
2	590	16.3	411	2	A04815	cellulase (EC 3.2.1.4	1.90e-100
3	566	15.7	515	2	S24943	endoglucanase - Clost	6.40e-96
4	660	15.6	814	1	CZCLEM	cellulase (EC 3.2.1.4	8.63e-95
5	614	14.5	364	2	S1207E	endoglucanase A - Rum	3.70e-86
6	603	14.2	409	2	S12018	endoglucanase B - Rum	4.21e-84
7	597	14.1	406	2	A43722	cellulase (EC 3.2.1.4	5.55e-83
8	528	12.5	482	2	JE0302	cellulase (EC 3.2.1.4	3.54e-70
9	525	12.4	473	2	S40507	endoglucanase - rumen	1.26e-69
10	510	12.0	517	2	I40798	cellulase (EC 3.2.1.4	7.29e-67
11	489	11.5	547	2	JQ0356	cellulase (EC 3.2.1.4	5.16e-63
12	479	11.3	475	1	CZCUGA	cellulase (EC 3.2.1.4	3.47e-61
13	470	11.1	455	2	S16559	cellulase (EC 3.2.1.4	1.52e-59
14	447	10.6	363	2	I40234	carboxymethylcellulas	2.27e-55
15	422	10.0	584	2	JQ1229	cellulase (EC 3.2.1.4	7.34e-51
16	412	9.7	900	2	JH0157	cellulase (EC 3.2.1.4	4.57e-49
17	324	7.7	500	2	S22458	cellulase (EC 3.2.1.4	1.44e-33
18	304	7.2	584	2	S27500	xylanase - Prevotella	3.88e-30
19	139	3.3	341	2	JC1201	carboxymethylcellulas	2.41e-04
20	141	3.3	418	1	S28372	cellulase (EC 3.2.1.4	1.31e-04
21	141	3.3	419	2	S72325	glucan 1,3-beta-glucos	1.31e-04
22	137	3.2	410	1	S68153	cellulase (EC 3.2.1.4	4.43e-04
23	126	3.0	338	2	JC4115	endoglucanase precurs	1.15e-02

[illegible]

ALTERNATE_NAMES endo-1,4-beta-glucanase E; endoglucanase E
 ORGANISM #formal_name Clostridium thermocellum
 DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Mar-1998

ACCESSIONS J0347
 REFERENCE J0347
 #authors Hall, J.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J.
 #journal Gene (1988) 59:29-38
 #title Conserved reiterated domains in Clostridium thermocellum endoglucanases are not essential for catalytic activity.

#cross-references MUID:89137992
 #accession J0347
 #molecule_type DNA
 #residues 1-814 #label HAL
 #cross-references GB:M27259; NID:9144768; PID:9144770
 #note part of this sequence, including the amino end of the mature protein, was confirmed by protein sequencing

COMMENT This secretory enzyme is part of a highly active and thermostable cellulase complex that is involved in extracellular cellulose degradation.

GENETICS
 #gene celE
 #start_codon GTG

FUNCTION catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in beta-D-glucans such as cellulose and lichenin; can hydrolyze such linkages in beta-D-glucans that also contain 1,3-linkages

#pathway cellulose degradation

CLASSIFICATION #superfamily cellulase CCA; Clostridium cellulase repeat homology

KEYWORDS duplication; extracellular protein; glycosidase; hydrolase; polysaccharide degradation

FEATURE
 1-34 #domain signal sequence #status predicted #label SIG
 35-814 #product cellulase E #status predicted #label MAT
 415-438 #domain Clostridium cellulase repeat homology #label CCR1
 451-474 #domain Clostridium cellulase repeat homology #label CCR2

SUMMARY #length 814 #molecular-weight 90244 #checksum 3788

Query Match 15.6%; Score 660; DB 1; Length 814;
 Best Local Similarity 37.1%; Pred. No. 8.63e-95;
 Matches 104; Conservative 72; Mismatches 92; Indels 12; Gaps 10;

Db 61 VKKIGKWNLTGDLA--PRTANGNPRRTKAMIEKREMGFNARVPTWDTHTGAPD 118
 QY 39 VRDMPGNLGNFTDVGQDETAGNPRVTRIELIERIADEGYSIRIPVTWENRIGGAPD 98

Db 119 YKIDEAWLNREVEVNVYVLCGVVAILNLHNDN-TWIIPTVANEQSKELVKVWEQIAT 177
 QY 99 YPIDPQFLNRVDEVQWALEEDLYVNLNLHSDSWLYEYEMHNTNGVMAKYSRLEWQSLN 158

Db 178 RKDYDDHLLFETMNEPREYGSMPWNGGTYENRD-VINRFLNAVVTIRASGNDKRF 236
 QY 159 HFKDYPTKLFESVNEPK-F-SQ-NW-GEIRENHALLDNLNTVEFVIRSGSQNDIRP 214

Db 237 ILVPTNAATGLDALNDL--VIPN-NDSRVIVSIHAYSPFFAMDVNGTSVWGSDDYKAS 293
 QY 215 LVLPMTMETATSQPLLNNLYQTDIKLDDPNLIATVHYGFWPFSVNIAGYTRFEED-SKRE 273

Db 294 LTSELDIYNRFVKNRGAIVGEGTIDKNNLSRVAHAE 333
 QY 274 IIEFDRVHHTFVARGIPVVLGEGFLGDKHTGVIQOGE 313

RESULT 5
 ENTRY #type complete
 TITLE endoglucanase A - Ruminococcus albus
 ORGANISM #formal_name Ruminococcus albus
 DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 13-Sep-1998

ACCESSIONS S12017
 REFERENCE S12017
 #authors Poole, D.M.; Hazlewood, G.P.; Laurie, J.I.; Barker, P.J.; Gilbert, H.J.
 #journal Mol. Gen. Genet. (1990) 223:217-223
 #title Nucleotide sequence of the Ruminococcus albus SV3 endoglucanase genes celA and celB.

#cross-references MUID:91066833
 #accession S12017
 #molecule_type DNA
 #residues 1-409 #label POO
 #cross-references GB:X54932; NID:945965; PID:945966
 CLASSIFICATION #superfamily Ruminococcus albus cellulase
 SUMMARY #length 409 #molecular-weight 45523 #checksum 266

Query Match 14.2%; Score 603; DB 2; Length 409;
 Best Local Similarity 29.4%; Pred. No. 4.21e-84;
 Matches 104; Conservative 100; Mismatches 125; Indels 25; Gaps 21;

Db 55 IPVSOHTNDPMTVTSARDLVAKMSGNWNLGNTMDATGELESEISWLPKVTYTKRMD 114
 QY 19 VAPAVSSANEDVKTLDIQSYVRDMPGNLGNFTFAVQ--D-ETANGNPRV-TRE-LIE 73

ACCESSIONS S12017
 REFERENCE S12017
 #authors Poole, D.M.; Hazlewood, G.P.; Laurie, J.I.; Barker, P.J.; Gilbert, H.J.
 #journal Mol. Gen. Genet. (1990) 223:217-223
 #title Nucleotide sequence of the Ruminococcus albus SV3 endoglucanase genes celA and celB.

#cross-references MUID:91066833
 #accession S12017
 #molecule_type DNA
 #residues 1-364 #label POO
 #cross-references GB:X54931; NID:945963; PID:945964
 #note the sequence from Fig. 4 is inconsistent with that from Fig. 2 in lacking 4-Pro, 5-Asp, 6-Lys, and 128-Asp and in having 77-Asn, 84-Ser, 89-Gly, 90-Val, and 108-Ile

CLASSIFICATION #superfamily Ruminococcus albus cellulase
 SUMMARY #length 364 #molecular-weight 41218 #checksum 3743

Query Match 14.5%; Score 614; DB 2; Length 364;
 Best Local Similarity 35.8%; Pred. No. 3.70e-86;
 Matches 101; Conservative 68; Mismatches 94; Indels 19; Gaps 16;

Db 21 EYRDISAMELVGEMKTHNLGNSLDATCAPGNASFEVNGNPKTKEMIDAVYKGFDIR 80
 QY 29 DVKTLDIQSYVRDMPGNLGNFTDAVG-Q-D-ETANGNPRVTRIELIERIADEGYSIR 84

Db 81 IPVTWGGHVGDPDKIDDEWIARVOEVVYAYDDGAVIINSHHEEDW-RIPDNEHI-D 138
 QY 85 IPVTWENRIGGAPDYPIDPQFLNRVDEVQWALEEDLYVNLNLHND-SWLYEYEMHNVN 143

Db 139 AVDEKTRAIKQVAERKDYGDHLIFEGLEPRVKVSGPQEWNGGTEGRRVC-DRLNKTF 197
 QY 144 GVMAKYRSLWEQLSNHFKDYPTKLFESVNEPKF--S-QNW-GEIRENHALLDNLTVF 199

Db 198 LDTVRATGNNNEKR-LLI--MTTYASSM-SNVIKDPAIPDDHIGFSIHATPYFTVN 253
 QY 200 FEIVRSGSQNDIRPLVLTMTATSQPLLNNLYQTDIKLDDPNLIATVHYGFWPFSVN 259

Db 254 ANADWELFHWDDSHDGEVLVSLMTNLKENYLDKIDPVIITEYG 295
 QY 260 I-AGYTRFE-EDSKR-EIIEFTDRVHHTFVARGIPVVLGEGF 298

RESULT 6
 ENTRY #type complete
 TITLE endoglucanase B - Ruminococcus albus
 ORGANISM #formal_name Ruminococcus albus
 DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 13-Sep-1998

ACCESSIONS S12018
 REFERENCE S12017
 #authors Poole, D.M.; Hazlewood, G.P.; Laurie, J.I.; Barker, P.J.; Gilbert, H.J.
 #journal Mol. Gen. Genet. (1990) 223:217-223
 #title Nucleotide sequence of the Ruminococcus albus SV3 endoglucanase genes celA and celB.

#cross-references MUID:91066833
 #accession S12018
 #molecule_type DNA
 #status preliminary
 #residues 1-409 #label POO
 #cross-references GB:X54932; NID:945965; PID:945966
 CLASSIFICATION #superfamily Ruminococcus albus cellulase
 SUMMARY #length 409 #molecular-weight 45523 #checksum 266

Query Match 14.2%; Score 603; DB 2; Length 409;
 Best Local Similarity 29.4%; Pred. No. 4.21e-84;
 Matches 104; Conservative 100; Mismatches 125; Indels 25; Gaps 21;

Db 55 IPVSOHTNDPMTVTSARDLVAKMSGNWNLGNTMDATGELESEISWLPKVTYTKRMD 114
 QY 19 VAPAVSSANEDVKTLDIQSYVRDMPGNLGNFTFAVQ--D-ETANGNPRV-TRE-LIE 73

```

Db 115 MLPEAGFNVLRIPIVSGNHLIDN-NYITDPAWMDRQEIIVNGYIDGDMYVILNTHHEW- 172
QY 74 RIADGKYKIRIPVTWENRIGGAPDYPIDPQFLNRVDEVQWALEEDLYVMINLHDSWL 133
Db 173 YM-PKPEKOGDIEELKAIWSQIADRFKGYDEHLIFEGLNPRLRGEGAEWTGTSEARE- 230
QY 134 WIYEMEHNYNGVMAKYSRLWEQLSNHFKDPTKLMFESVNEPKF-SQ--NWGEIRENHHA 190
Db 231 IINEYKAFVETVRASGNGNDRCIMI-T-GYAASSY-NNL-SAIELPEDSKLIISVH 286
QY 191 LDDDLNTVFFEIIVRSQSGQNDIRPLVPTMETATSQPLNNLYQTIDKLDOPN-LIAIVH 249
Db 287 AYLPSPALDTKTDKYPEDT-AIPLFESLNELFISRDIPVIVGEGSNKKNIDRV 345
QY 250 YYGFWPFSVNIAGYTRFEEDSKREIETFDVRVHTFVARGIPVVLGEGFLGDKHTGVI 309
Db 346 KCLDDYLGNAAKVDIPCVW-WDNVARI-GNGENFGLNRQEDYKFPKLMDFVK 397
QY 310 Q-OGKIL-KFFEYLIHNLNERDTHMLWDNGQHF---NRHTYEWDELFDMRLR 358

RESULT 7
ENTRY #type complete
TITLE cellulase (EC 3.2.1.4) precursor - Ruminococcus albus
ALTERNATE_NAMES endo-1,4-beta-glucanase
ORGANISM #formal_name Ruminococcus albus
DATE 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change
13-Sep-1998
ACCESSIONS A43722; A39134
REFERENCE A43722
#authors Ohmiya, K.; Kajino, T.; Kato, A.; Shimizu, S.
#journal J. Bacteriol. (1989) 171:6771-6775
#title Structure of a Ruminococcus albus endo-1,4-beta-glucanase
#accession A43722
#cross-references MUID:90078126
#status preliminary
#molecule_type DNA
#residues 1-406 #label OHM
#cross-references GB:M30928; NID:g152638; PID:g152639
REFERENCE A39134
#authors Ohmiya, K.; Deguchi, H.; Shimizu, S.
#journal J. Bacteriol. (1991) 173:636-641
#title Modification of the properties of a Ruminococcus albus
#accession A39134
#cross-references MUID:91100351
#status preliminary
#molecule_type DNA
#residues 44-70 #label OH2
ACTION
#description hydrolysis of 1,4-beta-D-glucosidic linkages in
beta-D-glucans such as cellulose and lichenin; can
hydrolyze such linkages in beta-D-glucans that also contain
1,3-linkages
#pathway cellulose degradation
CLASSIFICATION #superfamily Ruminococcus albus cellulase
KEYWORDS glycosidase; hydrolase; polysaccharide degradation
SUMMARY #length 406 #molecular-weight 45389 #checksum 2035

Query Match 14.1%; Score 597; DB 2; Length 406;
Best Local Similarity 31.6%; Pred. No. 5.55e-83;
Matches 92; Conservative 85; Mismatches 96; Indels 18; Gaps 14;

Db 53 VPVSQTHNTMTVTSKDLVAKWNTGNLGNMTDATAQGLGSEVSWPLKVTNKMID 112
QY 19 VAPAVSSANEDVKTLDIQSVQMDQPCWNLGNLTFDAVGQD--ETANGNPRVTR--LIE 73
Db 113 MLPEAGFNVLRIPIVSGNHIID-D-KYTSDFPAMWDRQEIIVNGYIDGDMYVILNTHHEW- 170
QY 74 RIADGKYKIRIPVTWENRIGGAPDYPIDPQFLNRVDEVQWALEEDLYVMINLHDSWL 133

```

```

Db 171 YM-PKPEKOGDIEELKAIWSQIADRFKGYDEHLIFEGLNPRLRGEGAEWTGTSEARE- 228
QY 134 WIYEMEHNYNGVMAKYSRLWEQLSNHFKDPTKLMFESVNEPKF-SQ--NWGEIRENHHA 190
Db 229 IINEYKAFVETVRASGNGNDRCIMI-T-GYAASSY-NNL-SAIELPEDSKLIISVH 284
QY 191 LDDDLNTVFFEIIVRSQSGQNDIRPLVPTMETATSQPLNNLYQTIDKLDOPN-LIAIVH 249
Db 285 AYLPSPALDTKTDKYPEDT-AIPLFESLNELFISRGIPVIVGEGFTM 334
QY 250 YYGFWPFSVNIAGYTRFEEDSKREIETFDVRVHTFVARGIPVVLGEGFL 300

RESULT 8
ENTRY #type complete
TITLE cellulase (EC 3.2.1.4) - rumen fungus (Neocallimastix
frontalis)
ALTERNATE_NAMES endo-1,4-beta-glucanase
ORGANISM #formal_name Neocallimastix frontalis
DATE 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change
05-Dec-1998
ACCESSIONS JE0302
REFERENCE JE0302
#authors Fujino, Y.; Ogata, K.; Nagamine, T.; Ushida, K.
#journal Biosci. Biotechnol. Biochem. (1998) 62:1795-1798
#title Cloning, sequencing, and expression of an endoglucanase gene
from the rumen anaerobic fungus Neocallimastix frontalis
MCH3.
#accession JE0302
#molecule_type mRNA
#residues 1-482 #label FUJ
#cross-references GB:U38843
GENETICS
#gene cels
KEYWORDS glycosidase; hydrolase
SUMMARY #length 482 #molecular-weight 54646 #checksum 7849

Query Match 12.5%; Score 528; DB 2; Length 482;
Best Local Similarity 30.8%; Pred. No. 3.54e-70;
Matches 88; Conservative 81; Mismatches 99; Indels 18; Gaps 15;

Db 78 ETCWGNPKTDMFKYLMNDQNFVFRPTTWSGHGFEAPDYKINEKWLKRVHEIVDYPK 137
QY 59 ETANGNPRVTRRELIERIADEGYKSRIPVTWENRIGGAPDYPIDPQFLNRVDEVQWALE 118
Db 138 NGAFVILNLHETWNHAFS-E-TLDTAKEILEKTSQAKFEKDYDEHLIFGGLNEPKN 195
QY 119 EDLYVMINLHDSWLWIYEMEHNTYNGVMAKYSRLWEQLSNHFKDPTKLMFESVNEP-KF 177
Db 196 DTPVETGGDQEGWDV-NAMNAVFLKTRSSGGNPKRHLMPPIYAAACNENSKN-F- 252
QY 178 -SQ-NW-GEIRENHALLDLDLNTVFFEIIVRSQSGQNDIRPLVPTMETATSQPLNNLYQ 234
Db 253 IFPE-DDDKVTASVHAYAPYNFALNNGAGAVDKFDAAGKDLKLEWINILMKRFRVDQIPM 311
QY 235 TIDKLDOPNLATVHYIGFWPFSVNI-AGYT-RFEEDSKREIETFDVRVHTFVARGIPV 292
Db 312 ILGEYGM-NRDN---EE-PRATWAEFYMEKVTAMGVPOVWONG 351
QY 293 VLGEFLGLGDKHTGVIQOQGEKLFEEYLIHHLNERDITHMLWDNG 338

RESULT 9
ENTRY #type complete
TITLE endoglucanase - rumen fungus (Neocallimastix patriciarum)
ORGANISM #formal_name Neocallimastix patriciarum
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
24-Jul-1997
ACCESSIONS S40507
REFERENCE S40507
#authors Zhou, L.; Xue, G.; Orpin, C.G.; Black, G.W.; Gilbert, H.J.;
Hazlewood, G.P.
#journal Biochem. J. (1994) 297:359-364

```


#title	Intronless celB from the anaerobic fungus Neocallimastix patriciarum encodes a modular family A endoglucanase.									
#accession	S40507									
#status	preliminary									
#molecule_type	mRNA									
#residues	1-473 ##label ZHO									
SUMMARY	#length 473 #molecular-weight 53070 #checksum 1241									
Query Match	12.4%; Score 525; DB 2; Length 473;									
Best Local Similarity	29.6%; Pred. No. 1,26e-69;									
Matches	104; Conservative 103; Mismatches 118; Indels 26; Gaps 20;									
Db	1	MKFLNTFLLS	AAIGSKAMKNISS--KELVKOLITIGWSLGTLDATCFETLDYKNQIA-	58						
Qy	1	MKKMWSMVLV	AVLVVSFVAPVSSANEDVKTLDIQ--SYVRDQMG-PW-LGNTEDAVGQ	57						
Db	59	SETCHGNVKT	TOELYKLSDLGNTFRIPRTTSGHFNAPDYKINDQMKRVRHEIVDYAI	118						
Qy	58	DETANGNPRVT	RELIERIADEGYSKIRIPVTWENRIGGAPDYDIPDQFLNRVDEVVQWAL	117						
Db	119	NTGGVAILNI	HETWNHAFQK--NLESAKKILVAIKWQATAAEFADYDEHILFEGMNEPRK	176						
Qy	118	EDLVYMLNLH	SDSLWLYMEHNYNGWAKYRSLEQSNHFKDPTKLMFESVNEP-K	176						
Db	177	VGDPAEWNGD	YEGWN-FYENMNDLVKTKIRATGGNNALRHLMPTVAAACINDGAINN-F	234						
Qy	177	FSQ--NW-GE	TRENHALLDDLTNTVFEIVROSGGQNDIRPLVPTWETATSOPLLNLY	233						
Db	235	K-FPSGDD-	KVIVLSHSYSPYNFALNAGAISNFYDGS--EIDWAMTINSKFISRGIP	290						
Qy	234	QTIDKLDPN	LNTATVHYIGFWFSPVNI-AG-YTRFEEDSKREIETFDVRVHTFVARGIP	291						
Db	291	VIIGFEGAM--	NRNN--ED-DRERWAERYIKKATSIGVPCVTWDNG-YPE	334						
Qy	292	VYLGEFLG	FDKHTGVIOQGEKLFEEYLIIHLNERDITHMLWDNGQHEN	342						
RESULT	10									
ENTRY	140798 #type complete									
TITLE	cellulase (EC 3.2.1.4) celA - Clostridium longisporum									
ALTERNATE_NAMES	endo-1,4-beta-D-glucanase									
ORGANISM	#formal_name Clostridium longisporum									
DATE	04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 21-Dec-1998									
ACCESSIONS	140798									
REFERENCE	140798									
#authors	Mittendorf, V.; Thomson, J.A.									
#journal	J. Gen. Microbiol. (1993) 139:3233-3242									
#title	Cloning of an endo-(1-->4)-beta-glucanase gene, celA, from the rumen bacterium Clostridium sp. ('C. longisporum') and characterization of its product, CelA, in Escherichia coli.									
#cross-references	MUID:94172316									
#accession	140798									
#status	preliminary; translated from GB/EMBL/DBJ									
#molecule_type	DNA									
#residues	1-517 ##label RES									
##cross-references	GB:L02868; NID:g144754; PID:g144755									
GENETICS										
#gene	celA									
CLASSIFICATION	#superfamily endoglucanase; bacterial cellulose-binding domain homology									
KEYWORDS	extracellular protein; glycosidase; hydrolase; polysaccharide degradation									
SUMMARY	#length 517 #molecular-weight 57660 #checksum 8670									
Query Match	12.0%; Score 510; DB 2; Length 517;									
Best Local Similarity	31.8%; Pred. No. 7,29e-67;									
Matches	91; Conservative 85; Mismatches 89; Indels 21; Gaps 17;									
Db	73	ETGWNPVTTK	AMDKIKNAGFKTIPIPTWGBHLDGN-N-KLNEEWKRVKEVDFCIA	130						
Qy	59	ETAWGNPRVT	RELIERIADEGYSKIRIPVTWENRIGGAPDYDIPDQFLNRVDEVVQWALE	118						

QY 284 TFVARGIPVVLGE 296

RESULT 12
ENTRY
TITLE C2CLCA #type complete
ALTERNATE_NAMES cellulase (EC 3.2.1.4) CCA precursor - Clostridium sp.
ORGANISM endo-1,4-beta-glucanase; endoglucanase-A
DATE #formal_name Clostridium sp.
30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 06-Dec-1996

ACCESSIONS JH0082
REFERENCE JH0082
#authors Faure, E.; Belaich, A.; Bagnara, C.; Gaudin, C.; Belaich, J.P.
#journal Gene (1989) 84:39-46
#title Sequence analysis of the Clostridium cellulolyticum endoglucanase-A-encoding gene, celCCA.
#cross-references MUID:90108715
#accession JH0082
#molecule_type DNA
#residues 1-475 #label FAU
#experimental_source ATCC 35319
#note a strong homology is found between the C-terminal duplicated region of this protein and regions in four endoglucanases and one xylanase from Clostridium thermocellum

GENETICS
#gene celCCA
#description catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in beta-D-glucans such as cellulose and lichenin; can hydrolyze such linkages in beta-D-glucans that also contain 1,3-linkages
#pathway cellulose degradation
CLASSIFICATION #superfamily cellulase CCA; Clostridium cellulase repeat homology
KEYWORDS duplication; extracellular protein; glycosidase; hydrolase; polysaccharide degradation
FEATURE
1-26 #domain signal sequence #status predicted #label SIG
26-475 #product cellulase #status predicted #label MAT
415-438 #domain Clostridium cellulase repeat homology #label CCR1
446-469 #domain Clostridium cellulase repeat homology #label CCR
SUMMARY #length 475 #molecular-weight 53624 #checksum 5839

Query Match 11.38; Score 479; DB 1; Length 475;
Best Local Similarity 30.78; Pred. No. 3.47e-61;
Matches 107; Conservative 96; Mismatches 110; Indels 35; Gaps 26;

33 PNLQIPQKNIPNNNDGMFVGLRGLGNLTGTFDAFNGTNTITNELDYETSWGKTTKOMI 92
QY 21 PAVSSANEDVKTLIDIOSYVRDMPGNLGNFTDA----V-GQ-D-ETANGNPRVTRELI 72
DB 93 DAIRKQGFNTVRIPVSWHPHVSQS-DYKISDVMMNRVQEVVNYCIDNKMXYILNTHHDVD 151
QY 73 ERTADEGYSIRIPVTWENRIGGAPDPIDPQFLNRVDEVVQWALEEDLYVMINLHHD-S 131
DB 152 KYGYPSPQYMASSKKYITVSWAQAARFANTYDEHLIFGCMNEPRVLGHAENWPELTN 211
QY 132 WLWIYEMEHNYNGVMARY-RSLQEQLSNHFKDPTKLMFESNEPKF-SQ-N--WGIRE 186
DB 212 SDVVDVINCINOLNODPVNTVRATGKNASRYLMCPGY-VASPDGATNDYFRMPNDISGN 270
QY 187 NH--HAL--LDDLTNTVFEIVRSGGQNDIRPLVLPMTETATSOPLNNLYQTIDKLD-D 241
DB 271 NKKIIVSHAYCPNFWAGLAMADGGTNAMNINSDQSEYTFWFMNDYNTKYSRGIPVII 330
QY 242 PN-LIATVHYGFWPFS-VNIA-GYTR-FE-EDSKREIET-F-DRVHTFVARGIPVL 294
DB 331 GEGGAV--DKNN--LK-T-RVEYMSYVAQAKARGILCIWDNN-NFS 371
QY 295 GFGLLGFDKHTGVIOOGEKLFKEFVLIHNLNERDITHMLWDNGQHPN 342

RESULT 13
ENTRY
TITLE S16559 #type complete
ORGANISM cellulase (EC 3.2.1.4) - Ruminococcus flavefaciens
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Mar-1998

ACCESSIONS S16559
REFERENCE S16559
#authors Cunningham, C.; McPherson, C.A.; Martin, J.; Harris, W.J.; Flint, H.J.
#journal Mol. Gen. Genet. (1991) 228:320-323
#title Sequence of a cellulase gene from the rumen anaerobe Ruminococcus flavefaciens 17.
#cross-references MUID:91360084
#accession S16559
#status preliminary
#molecule_type DNA
#residues 1-455 #label CUN
#cross-references GB:S55178; NID:g234871; PID:g234872
KEYWORDS glycosidase; hydrolase
SUMMARY #length 455 #molecular-weight 52420 #checksum 7708

Query Match 11.18; Score 470; DB 2; Length 455;
Best Local Similarity 29.48; Pred. No. 1.52e-59;
Matches 78; Conservative 68; Mismatches 102; Indels 17; Gaps 16;

DB 74 LDSAGLETGCGCPSEASQELFADAKAGFNTVRPTTWFOHLDEN-DN-IDPAWMARVH 131
QY 52 FDAVGQD-ETANGNPRVTRELIADGYSIRIPVTWENRIGGAPDPIDPFLNRVD 110
DB 132 QVVDYAYNIGLYIINLHHEQWNRADLATAYDDINPRLKMLTQIATEFKDYDHLIF 191
QY 111 EVQWALEEDLYVMINLHHD-SWLWIYEMEHNYNGVMARYSLWEQLSNHFKDPTKLMF 169
DB 192 ECMNEPRAMPTPWNWATPVVERDVINRLEANEVEIRGMDGPYAKTRLLMLPGYVASS 251
QY 170 ESNVEPK-F-SQ-NW-GEIRENHALLDLDLNTVFEIVRSGGQ-NDIRPLVLPMTETAT 224
DB 252 DKTFNLQIVLP-EN-DD-FVAVSIHATPYNTMTTEGAYHDTTKESNDLAYNLQ 308
QY 225 SQPLNNLYQTIDKLDPLNLIATVHYGFWPFSVNI-A--G-Y-TRFEEDSKREIETFD 279
DB 309 NFRDMFINKDIPVIGEMGTSDFCN 333
QY 280 RVHTFVARGIPVVLGEFLLGFDK 304

RESULT 14
ENTRY
TITLE I40234 #type complete
ORGANISM carboxymethylcellulase - Prevotella ruminicola
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 12-Aug-1996

ACCESSIONS I40234
REFERENCE I40234
#authors Matsushita, O.; Russell, J.B.; Wilson, D.B.
#journal J. Bacteriol. (1990) 172:3620-3630
#title Cloning and sequencing of Bacteroides ruminicola B-1-4 endoglucanase gene.
#cross-references MUID:90299778
#accession I40234
#status preliminary; translated from GB/EMBL/DBJ
#residues 1-363 #label RES
#cross-references GB:M38216; NID:g143940; PID:g143941
SUMMARY #length 363 #molecular-weight 40526 #checksum 7557

Query Match 10.68; Score 447; DB 2; Length 363;
Best Local Similarity 29.28; Pred. No. 2.27e-55;
Matches 79; Conservative 69; Mismatches 108; Indels 15; Gaps 11;

Db	137	IWEIAQREFKGSENLVEILNEP---H--GNITDQ---INDMKRIILNIIRKT---NP	185
Qy	152	LWEQLSNHFKDYPTKLMEFVNEPKFQSNWGEIRENHALLDDLNTVFEIVRQSGQND	211
Db	186	TRN-VI--IGAGWNSY-NLSQ-LETPDNPNIATPHYDYPSFTHQWG-TWGKNDM	239
Qy	212	IRPLVLTMTMETATSQLLNLYQITDKLDDPNLIATVHYGFWPFSVNIAGYRFEEDSK	271
Db	240	DAIAMVFNHYK-KWSDKNNPVILGEYGVMGHSDRSAY	277
Qy	272	REIETEDRHHHTVEARG-IPVVLGEFLGLGFKDKGTGI	309

Search completed: Fri Jun 4 10:05:29 1999
Job time : 76 secs.

```

RESULT      15
ENTRY
TITLE      cellulase (EC 3.2.1.4) CCD precursor - Clostridium sp.
ALTERNATE_NAMES
ORGANISM   #formal_name Clostridium sp.
DATE       31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
              22-Nov-1996
ACCESSIONS
REFERENCE   JQ1229
            Shima, S.; Igarashi, Y.; Kodama, T.
            Gene (1991) 104:33-38
            #authors
            #journal
            #title      Nucleotide sequence analysis of the endoglucanase-encoding
                        gene, celCCD, of Clostridium cellulolyticum.
            #cross-references MUID:92009193
            #accession   JQ1229
            #molecule_type DNA
            #residues    1-584 #label SHI
            #experimental_source AFCC 35319
            #note        the authors translated the codon ACG for residue 44 as
                        Phe and TTT for 45 as Thr

```

GENETICS	celCCD	fine and ill for 4.5 as ill
#gene		
FUNCTION		
#description	hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose and lichenin; can hydrolyze such linkages in beta-D-glucans that also contain 1,3-linkages	
#pathway	cellulose degradation	
CLASSIFICATION	#superfamily Clostridium cellulase repeat homology	
KEYWORDS	glycosidase; hydrolase; polysaccharide degradation	
FEATURE		
1-24	#domain signal sequence #status predicted #label SIG\	
25-584	#product cellulase CCD #status predicted #label CEL\	
329-353	#region proline/threonine-rich\	
530-553	#domain Clostridium cellulase repeat homology #label CCR1\	
562-584	#domain Clostridium cellulase repeat homology #label CCR2	
SUMMARY	#length 584	#molecular-weight 66061
		#checksum 2934

[illegible]